

Supplementary Information - Scannotation: a suspect screening tool for the rapid pre-annotation of the human LC-HRMS-based chemical exposome

Jade Chaker[‡], Erwann Gilles[‡], Christine Monfort, Cécile Chevrier, Sarah Lennon, Arthur David*

Univ Rennes, Inserm, EHESP, Irset (Institut de recherche en santé, environnement et travail) –

UMR_S 1085, F-35000 Rennes, France

[‡]Shared authorship (co-first authors)

*To whom correspondence should be addressed:

Tel: +33 299022885

email: arthur.david@ehesp.fr

Table of Content

Table of Content.....	1
1. Sample preparation.....	2
2. Data acquisition.....	2
3. Quality control procedures	3
4. Retention time prediction based on logP.....	4
5. Elements of proof to support the annotation of triclosan glucuronide.....	4

1. Sample preparation

Sample aliquots (200 μ L each) were spiked with a mix of 25 labeled internal standards (ISTD) at 100 ng/mL to monitor analytical variability during the sample preparation and data acquisition steps. A workup ultrapure water blank was also prepared with each sample preparation batch (n=5 batches of 25 samples each). Protein precipitation (PPT) was performed on all 75 samples using a 4:1 (v:v) ratio of cold methanol to matrix. Samples were then allowed to stand at -20°C for one hour prior to centrifugation to improve protein removal. After centrifugation at 13 300 rpm and 4°C for 20 min, supernatants were collected, homogenized, divided in two equal parts, and evaporated to dryness under vacuum. The first half was immediately resuspended in 40 μ L of injection phase (i.e., 90:10 (v:v) ultrapure water to acetonitrile ratio). This is further referred to as PPT extract. The second half was further purified through an additional extraction step. Firstly, samples were recovered in 400 μ L of a 75:25 (v:v) acetonitrile + 1% formic acid to ultrapure water ratio. Following homogenization and sonication, samples were cleaned up using a Phree (Phenomenex) protein and phospholipid removal plate drop by drop under vacuum. An additional volume of 100 μ L of the 99:1 (v:v) acetonitrile to formic acid mixture was drawn through the plate for rinsing. Phree extracts were then evaporated to dryness under vacuum, and recovered in 40 μ L of injection phase.

2. Data acquisition

Samples were analyzed on AB SCIEX X500R QTOF interfaced with an AB SCIEX ExionLC AD UPLC. Compound chromatographic separation was achieved using an Acquity UPLC HSS T3 C18 column (1.8 μ m, 1.0 x150mm) maintained at 40°C. Injection volume was set at 2 μ L. Flow rate was set at 100 μ L/min with mobile phases of ultrapure water/0.01% formic acid (A) and acetonitrile/0.01% formic acid (B). The gradient was as follows: 0-2.5 min, 10-20% B ; 2.5-20 min, 20-30% B ; 20-38 min, 30-45% B ; 38-45 min, 45-100% B ; 45-55 min, 100% B ; 55-60 min, 10% B. Full-scan mass spectra was acquired between 50-1100 m/z with a scan time of 0.328s, using ESI source settings: temperature 550°C, ionspray voltage 4,5kV (-4,5kV in negative mode), declustering potential 80V (-80V in negative mode),

accumulation time 300 ms, spray N₂ gas 35 arbitrary units, heat conduction gas 35 arbitrary units; curtain gas 7 arbitrary units, collisionally activated dissociation gas 7 arbitrary units, run time 60min. Samples were analyzed in full scan experiment in both – and + ESI modes. MS/MS mass fragmentation information for chemical elucidation was obtained by further analysis of selected samples in data dependent acquisition (DDA). DDA experiments were performed in both – and + ESI modes, using the following source settings: MS1 accumulation time 250ms, MS2 accumulation time 100ms, collision energy 35eV, cycle time 2.346s, mass range 50-1100 m/z. Precursor ion selection parameters were as follows: a maximum of 20 candidate ions per cycle, intensity threshold 1cps, and dynamic background subtraction was enabled (candidate ions only includes ions increasing in intensity).

3. Quality control procedures

One solvent blank (i.e., initial gradient composition) and one extraction blank sample (i.e., sample preparation with UHPLC grade water instead of sample) were systematically injected with each batch, to respectively ensure lack of carryover in the UHPLC system and to monitor the contamination linked to the sample preparation process. In particular, sample preparation-linked contamination for annotated compounds was taken into account by verifying their presence in the extraction blank, and if so, subtracting the blank area from the samples' areas. Batch assignment and injection order were randomized. Moreover, composite quality control samples were prepared and injected repeatedly (i.e., 6 times) after blanks to equilibrate the analytical system, and to periodically monitor throughout the batch (i.e., every 5 samples) the analytical drift and repeatability. Overall, 55 QCs were injected for each sample preparation type and ESI mode (i.e. 220 in total). Repeatability was assessed by evaluating the coefficient of variation of feature areas over the 55 QCs, which had to be under 30% for over 80% of features. Similarly, median area coefficient of variation was expected to be under 20%. ISTD were systematically used in samples to assess analytical drift and were manually integrated using vendor software Sciex OS v.1.6 (AB SCIEX). Median area CV of ISTDs in QC samples and in cohort samples was expected to be under 20%, and median Rt of ISTDs was expected to be under 15%.

4. Retention time (Rt) prediction based on logP

Regarding the logP-based Rt, a simple linear regression model was chosen as logP has been demonstrated as a good determinant of Rt¹⁻³. Here, the model parameters are not hard-coded into the software, as it is up to each user to import their own logP and experimental Rt values to create the linear regression that fits their hardware and analytical method. For our in-lab library, the model parameters were validated in previous work⁴ by training it on a –compound dataset and validating it on a 30-compound dataset. Users may provide logP and experimental Rt values for their suspects (at least 20 pairs) to train the model to their own analytical setup. The training compounds should be chosen to be diverse enough to be representative of the analyzed chemical space in terms of m/z, Rt, and chemical family. It should be noted that if the linear regression of user data yields an r^2 under 0.6, Scannotation does not calculate logP-based Rt, as it would consider the model to not be robust enough.

5. Elements of proof to support the annotation of triclosan glucuronide

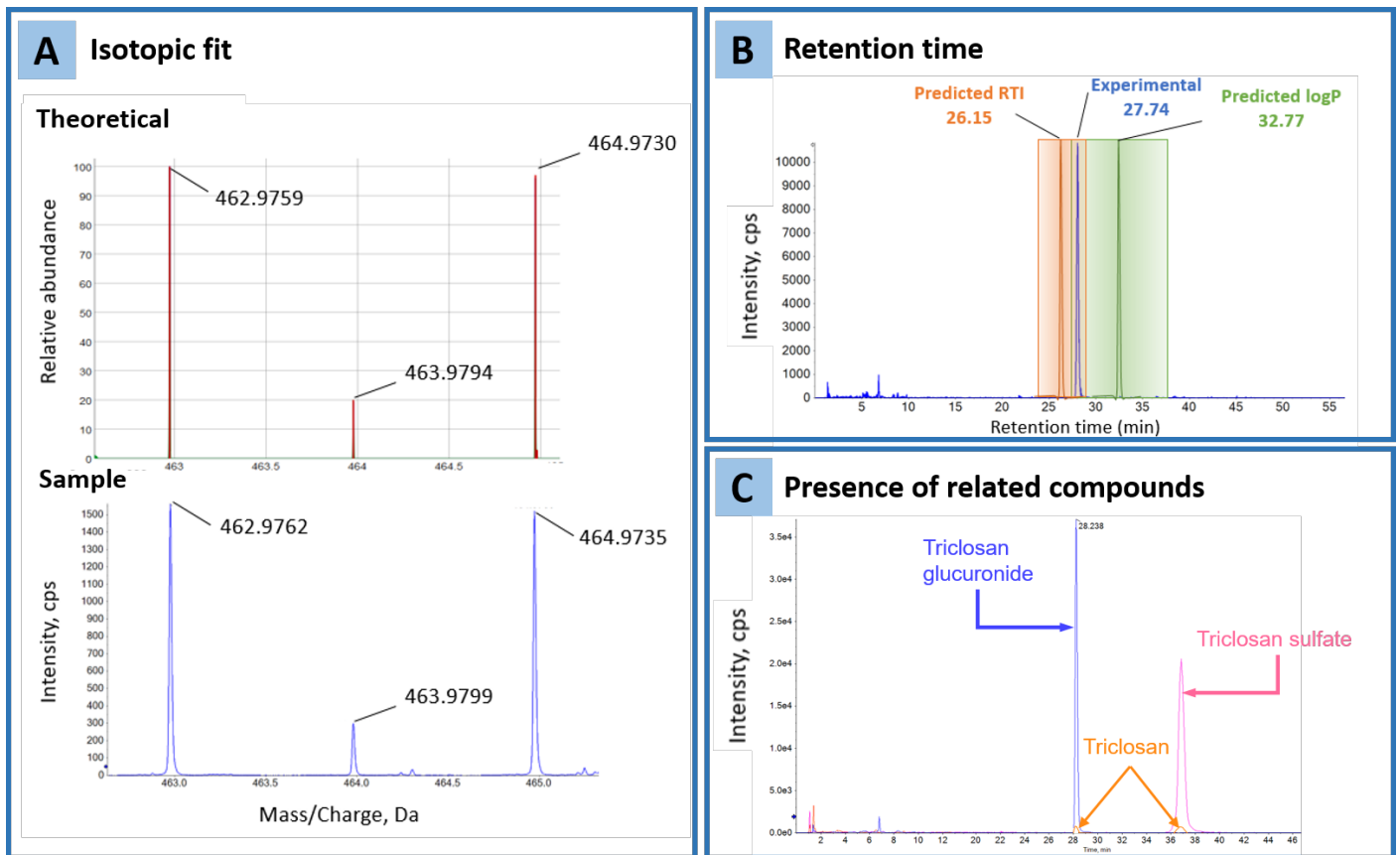


Figure S1 - Evidence supporting the annotation of triclosan glucuronide: coinciding theoretical and experimental isotopic profiles (A); experimental retention time coherent with different retention time predictions (B); and presence of other related compounds, such as triclosan sulfate, as well as observation of neutral losses coherent with it being a triclosan-derived compound (C).