



**HAL**  
open science

## Non-targeted screening methodology to characterise human internal chemical exposure: Application to halogenated compounds in human milk

Mariane Pourchet, Luca Narduzzi, Annabelle Jean, Ingrid Guiffard, Emmanuelle Bichon, Ronan Cariou, Yann Guitton, Sébastien Hutinet, Jelle Vlaanderen, Jeroen Meijer, et al.

### ► To cite this version:

Mariane Pourchet, Luca Narduzzi, Annabelle Jean, Ingrid Guiffard, Emmanuelle Bichon, et al.. Non-targeted screening methodology to characterise human internal chemical exposure: Application to halogenated compounds in human milk. *Talanta*, 2021, 225, pp.1-10. 10.1016/j.talanta.2020.121979 . hal-03338400

**HAL Id: hal-03338400**

**<https://hal.inrae.fr/hal-03338400v1>**

Submitted on 16 Dec 2022

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial 4.0 International License

1            Non-targeted screening methodology to characterise human internal chemical  
2                            exposure: application to halogenated compounds in human milk

3

4    Mariane Pourchet<sup>1</sup>, Luca Narduzzi<sup>1</sup>, Annabelle Jean<sup>1</sup>, Ingrid Guiffard<sup>1</sup>, Emmanuelle Bichon<sup>1</sup>, Ronan Cariou<sup>1</sup>,  
5    Yann Guitton<sup>1</sup>, Sébastien Hutinet<sup>1</sup>, Jelle Vlaanderen<sup>2</sup>, Jeroen Meijer<sup>2</sup>, Bruno Le Bizec<sup>1</sup>, Jean-Philippe  
6    Antignac<sup>1\*</sup>

7

8    <sup>1</sup>Oniris, INRAE, LABERCA, 44307 Nantes, France

9    <sup>2</sup>Institute for Risk Assessment Sciences (IRAS), Utrecht University, Utrecht, The Netherlands

10

11    **Keywords:** Non-targeted screening, human milk, human biomonitoring, internal chemical exposure, HRMS,  
12    HBM4EU

13    \* **Corresponding author:** Jean-Philippe Antignac, Laboratoire d'Etude des Résidus et Contaminants dans  
14    les Aliments (LABERCA), Route de Gachet, CS 50707, Nantes, F-44307, France, [laberca@oniris-nantes.fr](mailto:laberca@oniris-nantes.fr)

15

16 **Abstract**

17 Suspect and non-targeted screening approaches are a matter of increasing interest notably with regard to the  
18 Exposome contextual framework, but their application to human samples still remains limited at this date.  
19 The aim of the present study was to develop a non-targeted workflow from sample preparation to data  
20 processing and method assessment to characterise the human internal chemical exposure at early life stage.  
21 The method was focused on human milk to investigate mother and newborn exposure to known organic  
22 contaminants and to extend the characterisation to unknown compounds. We specifically focused on  
23 halogenated biomarkers of exposure due to persistence and potential toxicological impact reasons. The  
24 newly developed approach was based on a simple and fast sample preparation followed by a comprehensive  
25 analysis by both liquid and gas phase chromatography coupled to high resolution mass spectrometry. Critical  
26 steps of the non-targeted workflow as the method assessment have been addressed with a reference mix of 30  
27 chlorinated and brominated contaminants encompassing various substances groups and a statistical approach.  
28 Data processing until the identification of biomarkers of exposure was possible with homemade  
29 bioinformatics tools. On the other hand, the method was validated by the identification of historical  
30 chemicals as hexachlorobenzene and p,p'-DDE and emerging chemical as 4-hydroxychlorothalonil. This  
31 approach opens the door to further extensions and consolidations to offer new capabilities for exposomics  
32 and environmental health research.

33

34

35

36

37

38

39

40

## 41 Introduction

42 Humans are exposed to thousands of chemicals in the environment, most of which are likely unknown.  
43 These compounds comprise our exposome, as defined by C.P. Wild in 2005 [1]. The impact of the exposome  
44 on human health has been a matter of increasing concern in recent years [2-4]. Human biomonitoring (HBM)  
45 programmes are conducted to assess current human exposures to environmental compounds and to depict  
46 temporal trends reflecting changes of patterns and levels of exposure [5] as well as to measure the effect and  
47 efficiency of policy regulations. Among these programmes is the Human Biomonitoring for Europe H2020  
48 project (HBM4EU, 2017-2021) which aims to characterise the chemical exposure of European populations,  
49 to provide evidence of possible environmental health effects and to support public policy making [6].

50 Although humans are exposed to environmental chemical contaminants throughout their entire lives, certain  
51 periods of exposure are especially critical. The World Health Organisation identifies childhood (*i.e.* from  
52 conception to adolescence, pregnancy and nursing [7]) as an exposure window of interest. This holds  
53 significance not only for the direct future health of children who are exposed today, but also for the indirect  
54 health of their potential future offspring or even their future ability to sire offspring. This generational  
55 exposure transfer is of high concern particularly for foetuses/newborns exposed during pregnancy via cord-  
56 blood and placenta transfer [8] and during the first year of life via breastfeeding [9, 10] because their blood-  
57 brain barriers are not yet fully developed. Chemical exposures during development stages can increase the  
58 risk of potentially irreversible damage, such as cognitive and motor development issues associated with  
59 perinatal exposure to some persistent organic pollutants such as PCBs [11,12]. Despite this, evidence  
60 supports the role of breastfeeding in newborn immunologic protection, as human milk composition adapts to  
61 the baby's growth over the course of lactation [13]. Consequently, this matrix is of particular interest for  
62 further investigation with regard to these chemical exposure-health relationships. Human milk is, however, a  
63 challenging biological matrix for analytical scientists due to its hydrophilic and lipophilic properties (fat  
64 content of 2 to 4%) [14], and requires particular attention for sample preparation and extract compatibility  
65 for analysis.

66 Matrix specificity and related challenges have been historically well addressed with targeted methods. These  
67 approaches are developed to detect and quantify known chemicals. Targeted methods focus on a set of

68 contaminants with known physicochemical properties [15,16]. In order to extend the range of detected  
69 markers of exposure and to generate a broader picture of the human chemical exposome, there is increasing  
70 need to develop non-targeted screening (NTS) approaches. This trend can already be seen applied to various  
71 fields such as the environment [17-20] and food [21,22]. Using these large scale NTS approaches is  
72 particularly promising for detecting chemicals of emerging concern (CECs) as early warning signs for risk  
73 assessment. However, these new methodological approaches still face a number of limitations and challenges  
74 [23]. First the sample preparation should reach a compromise between sufficient selectivity to remove matrix  
75 interfering compounds (for instance proteins and lipids for human milk), while preserving a maximum of  
76 other compounds to assure a large screening. Then, the analysis should cover a wide range markers of  
77 exposure by using complementary techniques as liquid and gas phase chromatography (LC and GC) coupled  
78 to high resolution mass spectrometry (HRMS).The high dimensional data typically generated by these  
79 approaches should be processed with appropriate bioinformatics tools. The goal of NTS approaches is not to  
80 exhaustively characterise the exposome but rather to prioritise compounds to investigate, as for instance  
81 focus on halogenated compounds [24] already known for hazard effect.

82 One component of the HBM4EU project aims to develop non-targeted methods applied to human matrices in  
83 order to characterise citizen exposure to CECs. In that context, the present paper describes the development  
84 of methods applied to the NTS of halogenated biomarkers of chemical exposure (measured by the presence of  
85 exogenous substances or their metabolites) in human milk from sample preparation to data generation and  
86 processing. In particular, sample preparation and LC/GC-HRMS analysis aimed to cover a broad range of  
87 molecules, with an expected compromise between high-level selectivity and *ad hoc* sensitivity. HaloSeeker  
88 [25], was used to focus the data processing on chlorinated and brominated contaminants. In addition, the  
89 present article presents an innovative approach to assess the method's performance by using an Orthogonal  
90 Partial Least Squares (OPLS) model. . Based on this multivariate regression model the signal intensities  
91 obtained for a range of reference standard compounds was related to their individual physicochemical  
92 properties. This predictive model was named qsRecr for quantitative structure-recovery relationship, as an  
93 image of the first work published by Kaliszan in 1992 [26] dealing with the quantitative structure-retention

94 relationship (qsrr). Finally, a proof-of-concept intending to demonstrate the suitability of the developed  
95 approach was conducted on a composite of human milk.

## 96 **1 Material and method**

### 97 1.1 Chemicals and reagents

98 All chemicals and solvents used were of high quality grade for trace analysis. Acetonitrile, water, acetone  
99 and isopropanol were obtained from Sigma-Aldrich (LC-MS ChromaSolv grade, St. Louis, MO, USA).  
100 Ammonium acetate salt (Emsure grade) was purchased from Merck (Darmstadt, Germany). Hexane was  
101 purchased from LGC Promochem® (Picograde quality Wesel, Germany). Captiva EMR-Lipid (enhanced  
102 matrix removal of lipids) 6 mL, 600 mg cartridges were obtained from Agilent Technologies (USA).

### 103 1.2 Standards

104 A solution of 30 halogenated compounds (Table S1), was constituted to assess method performance and  
105 named “QA/QC compounds mix”. It included acetochlor, beta-hexachlorocyclohexane ( $\beta$ -HCH), alpha-  
106 hexabromocyclododecane ( $\alpha$ -HBCDD), triclosan, 6-hydroxy-2,2',3,4,4',5-hexabromodiphenyl ether (OH-  
107 BDE 137), 2,2',4,4',5,5'-hexabromodiphenyl ether (PBDE 153), quizalofop-p-ethyl, 1,2-bis(2,4,6-  
108 tribromophenoxy) ethane (BTBPE), hexabromobenzene (HBBz), 3,5,6-trichloro-2-pyridinol (TCPy), anti-  
109 dechlorane plus (anti-DP), purchased from AccuStandard Inc. (New Haven, CT, USA). Deltamethrin and  
110 dichlorodiphenyltrichloroethane (p,p'-DDE) were purchased from Agilent (North Kingstown, Rhode Island,  
111 U.S.). Simazine, hexachlorobenzene (HCB), fipronil, chlorpyrifos, 2,3,4,5-tetrachlorophenol (2,4,3,5-  
112 tetraCP), prochloraz, (Z)-dimethomorph, fenhexamid, fenvalerate free acid, chlorfenvinphos, metolachlor,  
113 2,4-dichlorophenol (2,4-DCP), 2,4-dibromophenol (2,4-DBP) and tetraconazole were purchased from  
114 Dr Ehrenstorfer GmbH (Augsburg, Germany). 2,3,4,5-tetrabromophenol (2,4,3,5-tetraBP) was purchased  
115 from Cambridge Isotope Laboratories (Andover, USA) and tetrabromobisphenol A (TBBPA) from Acros  
116 organics. Compounds were diluted in two stock solutions (1 ng  $\mu\text{L}^{-1}$ ), in acetonitrile for standard received in  
117 methanol or acetonitrile, and in toluene for standards received in toluene, cyclohexane or isooctane. Then,  
118 both solutions were combined and diluted in acetonitrile (0.1 ng  $\mu\text{L}^{-1}$ ). The multivariate regression model  
119 was assessed with three analytical standards of atrazine, purchased from Dr Ehrenstorfer GmbH and

120 perfluorooctane sulfonate ( $^{13}\text{C}_4$ -PFOS) and bis(1,3-dichloro-2-propyl) phosphate ( $^2\text{H}_{10}$ -BDCIPP) from  
121 Wellington Laboratories (Guelph, Ontario, Canada). 4-Hydroxychlorothalonil was purchased from  
122 Dr Ehrenstorfer GmbH. Quality control of the analytical workflow was assured with internal labelled  
123 standards ( $^2\text{H}_{18}$ - $\alpha$ -HBCDD,  $^{13}\text{C}_{12}$ -TBBPA and  $^{13}\text{C}_6$ -HBBz) and external labelled standards ( $^2\text{H}_{18}$ - $\gamma$ -HBCDD  
124 and  $^{13}\text{C}_3$ -atrazine for fractions analysed by LC and  $^{13}\text{C}_{10}$ -anti-DP for fractions analysed by GC), at 0.1 ng  $\mu\text{L}^{-1}$ .  
125 They were all obtained from Wellington Laboratories, except  $^{13}\text{C}_{10}$ -anti-DP which was purchased from  
126 Cambridge Isotope Laboratories.

### 127 1.3 Samples

128 Human milk samples used for the present method development were originated from a French mother-child  
129 study described elsewhere [27,28]. They were collected in 2008 and stored at -20 °C until analysis. Five  
130 samples were thawed, vigorously agitated and pooled (2.5 mL each). This composite human milk sample  
131 was used to assess the method performance and as a proof-of-concept.

### 132 1.4 Sample preparation

#### 133 1.4.1 Protocol

134 Sample preparation was intended to be as non-selective as possible to preserve markers of exposure while  
135 removing matrix interferents (Figure 1). Human milk or water (procedural blank) aliquots (400  $\mu\text{L}$  each)  
136 were mixed with 1.6 mL of acetonitrile containing standards (internal standards and QA/QC compounds mix  
137 for method assessment). After protein precipitation, samples were centrifuged (10 min at 4°C, 2500 g) and  
138 purified on Captiva EMR-Lipid cartridge, chosen for its polymeric phase selective of lipids by size exclusion  
139 and hydrophobic interactions [29-31]. Supernatants were loaded on Captiva EMR-Lipid cartridge previously  
140 conditioned with 10 mL of acetonitrile/water 80:20 (v/v) and eluted at atmospheric pressure without adding  
141 any further eluent. Eluate was extracted with 2x2 mL of hexane. Both fractions were concentrated until  
142 dryness under a gentle nitrogen stream at 35 °C for acetonitrile/water extract and at room temperature for  
143 hexane extract. Final extracts were re-suspended in 50  $\mu\text{L}$  of acetonitrile/water 80:20 (v/v) or 50  $\mu\text{L}$  of  
144 hexane, both containing *ad hoc* external standards for LC- and GC-HRMS analysis, respectively.

## 145 1.5 HRMS measurement methods

### 146 1.5.1 LC-ESI(+/-)-Q-Orbitrap

147 The present LC-HRMS instrumental method was adapted and modified from a previously developed [22].  
148 Sample extracts (5  $\mu$ L) were injected onto a Hypersil Gold column (100 mm  $\times$  2.1 mm, 1.9  $\mu$ m)  
149 (ThermoFisher Scientific, San José, CA, USA) kept at 45  $^{\circ}$ C and controlled by an UltiMate 3000 UHPLC. A  
150 gradient with a flow rate of 0.4 mL min<sup>-1</sup> was run using water (A), acetonitrile (B), both supplied with 10 mM  
151 of ammonium acetate, and isopropanol/acetone 1:1 (v/v) (C). The gradient began with A/B 80:20 (v/v) for 2  
152 min, ramped to 0:100 in 16 min and hold for 3 min. It ramped to B/C 10:90 for 2 min, hold for 5 min,  
153 returned to 100% B for 1 min and held for 2 min before returning to initial conditions in 1 min and stabilised  
154 for 10 min. The LC system was coupled to a Q-Exactive<sup>TM</sup> mass spectrometer through a heated electrospray  
155 ion source (HESI-II, ThermoFisher Scientific). External mass detector calibration was performed before each  
156 batch by infusing calibration mixture for negative and positive ionisation mode (MSCAL6 and MSCAL5  
157 ProteoMass LTQ/FT-Hybrid, Supelco, Bellefonte, PA, USA). Data were acquired in full scan mode over the  
158 mass-to-charge ratio ( $m/z$ ) range 100–1000 at a resolving power of 140,000 full width half maximum  
159 (FWHM) at  $m/z$  200. Automatic gain control (AGC Target) was set at high dynamic range ( $5 \times 10^5$ ) and  
160 maximum injection time (IT) at 500 ms. For both ionisation modes, ran in separate injections, parameters  
161 were: sheath gas flow, 50 arbitrary units (AU); auxiliary gas flow, 10 AU; capillary temperature, 350  $^{\circ}$ C;  
162 heater temperature, 350  $^{\circ}$ C; S-lens radio frequency, 70 AU. Spray voltage was set at 3.5 kV in positive mode  
163 and -2.5 kV in negative ion mode. Instrument was controlled by Xcalibur (ThermoFisher Scientific) software  
164 version 3.0.

### 165 1.5.2 GC-EI-Q-Orbitrap

166 The GC-HRMS instrumental method was developed on the basis of general settings originated from previous  
167 expertise aggregated from targeted methods dedicated to various compound classes. Samples extracts (2  $\mu$ L)  
168 were injected in splitless mode onto a DB5-MS column (30 m  $\times$  0.25 mm, 0.25  $\mu$ m) (Agilent, Palo Alto, CA,  
169 USA). The GC oven temperature was programmed as follows: initial temperature was set to 60  $^{\circ}$ C for 2 min  
170 and increased to 130  $^{\circ}$ C at 10  $^{\circ}$ C/min, then to 250  $^{\circ}$ C at 5  $^{\circ}$ C/min and to 320  $^{\circ}$ C at 10  $^{\circ}$ C/min and held at 320  
171  $^{\circ}$ C for 10 min. The total run time was 50 min. Injector temperature was set to 320  $^{\circ}$ C and flow rate to 1 mL



172 min<sup>-1</sup> helium, constant flow. Compounds were conducted through a transfer line heated at 320 °C to the  
173 electron impact ion source heated at 300 °C, with an electronic energy set at 70 eV. The solvent delay was  
174 set to 5 min. Data was acquired in full scan mode over the *m/z* range 50–750 at a resolving power of 120,000  
175 FWHM at *m/z* 200. AGC Target was set at 5×10<sup>5</sup> ions with automatic filling limit and maximum IT at 500  
176 ms. Instrument was controlled by Xcalibur software version 4.1.

## 177 1.6 Data processing

178 For the method assessment, Xcalibur was used to integrate chromatographic peaks, verify retention time and  
179 mass spectra of QA/QC compounds mix, with a mass tolerance of 5 ppm.

180 For the proof-of concept, LC data were processed using HaloSeeker v1.0. Peak-picking parameters were *m/z*  
181 tolerance, 5 ppm; peakwidth, 5-60 s; pre-filter step, 3; pre-filter level 10 000 and sntresh, 10. Halogen-  
182 pairing parameters were *rT* tolerance, 1 second and *m/z* tolerance, 0.5 mDa. Additional tools (in  
183 development to be included in the next version of HaloSeeker) aligned samples with a bandwidth of 1 s. For  
184 blank subtraction, all signals found at least once in a procedural blank sample and in real samples were not  
185 taken into account. Monoisotopic masses ion only detected in milk were matched with the HBM4EU list.  
186 This list was elaborated in the frame of the HBM4EU project and it merges several databases (NORMAN,  
187 EWAG, etc.) resulting in a list of approximately 70 000 compounds. GC data processing is more challenging  
188 mainly because GC-EI-MS spectral database are unit based (*e.g.* NIST) [32,33] and because of high  
189 fragmentation, so more than a list, a real database including fragmentation spectra acquired in HRMS is  
190 required for identification. We recently started the elaboration of such homemade database. However, it is  
191 not large enough to cover thousands of compounds. Now, it mainly included contaminants from the QA/QC  
192 compound mix. In this context, GC data were manually processed by extraction of ion chromatograms with  
193 Xcalibur with a mass tolerance set at 5 ppm. For both LC and GC, compound detection in all triplicate was  
194 required for further consideration.

## 195 1.7 Method performance assessment

### 196 1.7.1 Classical criteria

197 Method efficiency was assessed based on five classical criteria including sensitivity, calibration,  
198 repeatability, recovery and matrix effect (ME), using the QA/QC compounds mix. Contaminants were  
199 selected regarding their physicochemical properties and associated coverage of an extended range of  
200 potential markers of interest representative of external and internal exposure (native contaminants and  
201 metabolites): molecular size (160 to 690 Da), polarity (log P from 2.2 to 8), chlorinated and brominated  
202 degree and detectability in GC-EI and/or LC-ESI(+) and/or LC-ESI(-) (Figure 2).

203 Instrumental limit of detection (LOD) was assessed with an external calibration curve of QA/QC compounds  
204 mix at concentration 0.001; 0.005; 0.01; 0.05; 0.1 and 0.5 ng  $\mu\text{L}^{-1}$ . The LOD was considered as the lowest  
205 concentration leading to the detection of a chromatographic peak defined as at least 5 consecutive scans and  
206 S/N higher than 3 at expected retention time in LC and GC.

207 The linear calibration curve of four concentration levels (6.25; 12.5; 62.5; 125 pg  $\mu\text{L}^{-1}$  of human milk) was  
208 accepted if the coefficient of determination ( $R^2$ ) was greater than 0.99.

209 Repeatability was assessed as the relative standard deviation (RSD) of the signal intensities observed for  
210 each compound and each concentration level in triplicate. The repeatability was considered acceptable if  
211  $\text{RSD} < 30\%$ .

212 Recovery was calculated with equation (1). It took into account matrix effects and loss occurring during both  
213 sample preparation and ionisation. Recovery was calculated for all concentration levels with the instrumental  
214 method offering the lowest LOD. However, the liquid-liquid partitioning may split compounds in two  
215 fractions leading to specific recoveries, measured independently in LC and in GC. In this case, if recoveries  
216 were higher than 10 % in both fractions, the total recovery (sum of both) was considered. Recovery yields  
217 were deemed acceptable for screening purpose if higher than 15%. Below this value the corresponding  
218 compound was considered out of the method application range.

219 ME was calculated according to equation (2) at the concentration level 62.5 pg  $\mu\text{L}^{-1}$  human milk. The mean  
220 of ME observed for each sample prepared in triplicate is reported. The result has been evaluated by expert  
221 opinion, taking into account its consequence with regard to the repeatability and sensitivity.

$$222 \quad \text{Equation 1: Recovery (\%)} = \left( \frac{\text{Area}_{\text{comp.spiked be. ext.}} - \text{Area}_{\text{comp.in unspiked sample}}}{\text{Area}_{\text{Standard}}} \right) * 100$$

$$223 \quad \text{Equation 2: Matrix effect (\%)} = \left( \frac{\text{Area}_{\text{Comp. spiked af. ext.}} - \text{Area}_{\text{comp.in unspiked sample}}}{\text{Area}_{\text{Standard}}} - 1 \right) * 100$$

224 Where “comp” means “compound”, “be. ext.” is “before extraction” and “af. ext.” is “after extraction”.

### 225 1.7.2 Predicting the recovery through the physicochemical properties of the considered markers

226 Recovery can be measured using the standards addition method. Unfortunately, not all analytical standards  
227 are commercially available or accessible for cost or other practical reasons, especially for new or unknown  
228 compounds sought by large scale non-targeted screening approach. It is well-known that main factors  
229 influencing the recovery of a given compound are related to its physicochemical properties; a multivariate  
230 regression model based on Orthogonal-Partial Least Squares (OPLS) approach [34] was adopted to better  
231 characterise method limitations and efficiency. The model was built using SIMCA-P 13.0 (Umetrics, Umea,  
232 Sweden). It evaluates whether compounds' recovery (matrix Y) could be predicted through their chemical  
233 structures (matrix X). The model was built with the 30 reference standard compounds characterised by 23  
234 different physicochemical parameters extracted from PubChem, including exact mass, mass defect, relative  
235 mass defect (RMD, [35]), topological polar surface area, log P, complexity, number of each atoms (C, H, N,  
236 O, Br, Cl, P, F, S), number of unsaturation, presence of OH group or oxygen double bond, presence of OH  
237 group, non-ramified cycle, heavy atom count, hydrogen bond donor count, hydrogen bond acceptor count,  
238 rotatable bond count (table S2). The descriptors of the compounds' structures have been regressed versus the  
239 measured recovery from LC or GC data, or the sum of both for specific cases (see detail section 2.7.1.), for  
240 the concentration level 62.5 pg  $\mu\text{L}^{-1}$  of human milk. Due to possible synergic effects, the squared and crossed  
241 descriptor terms were also included in the model. The obtained model was validated using *k*-fold cross-  
242 validation and considered acceptable when  $R^2$  and  $Q^2$  indicators were above 0.5, with an associated *p*-value  
243 below 0.05. The model accuracy was estimated as root mean squared error (RMSE) [36].

## 244 1.8 Proof-of-concept design

245 Unspiked composite of human milk and procedural blank samples were prepared in triplicate using the same  
246 method in order to determine method compatibility for non-targeted screening. Samples were spiked with  
247 labelled internal and external standards.

## 248 2 Results and discussion

### 249 2.1 Method performance

250 First, the LC versus GC chromatographic complementarity was evaluated. The GC-EI-HRMS method  
251 detected 23 out of the 30 reference compounds included in the QA/QC compounds mix. However, some  
252 compounds were more efficiently detected in LC-ESI(+/-)-HRMS with a much lower limit of detection  
253 (LOD) (*e.g.* phenolic molecules) (Table S3). The ideal situation would be to analyse each fraction with both  
254 detection modes, leading to a total of six injections per sample, approximately 4.5 hours of analysis, and  
255 laborious data processing. This is feasible for few samples but not realistic in a context of high throughput  
256 exposomics research to support large scale biomonitoring studies.

257 For each considered exposure marker, the detection mode leading to the lowest LOD was selected for further  
258 investigation. As an exception, acetochlor and metholachlor exhibited similar LOD both in LC-ESI(+) and  
259 GC-EI. As the estimated recovery was found higher than 10 % in both fractions for these two compounds,  
260 the sum of the two values was calculated and considered for further interpretation. Conversely, LODs  
261 observed for chlorfenvinphos in GC-EI and LC-ESI were found similar but the recovery calculated for the  
262 GC fraction was lower than 10 %. Only the LC-ESI(+) signal was then considered for this last compound  
263 for further investigation. Lastly, tetraconazole was detected in three of the investigated instrumental  
264 methods; the lowest LOD was obtained in LC-ESI(-) ( $<0.001 \text{ ng } \mu\text{L}^{-1}$ ). The different recoveries observed for  
265 all the considered markers are summarised in Table 1.

266 Repeatability was globally satisfactory with RSD lower than the fixed limit of 30% for most compounds  
267 (Table S4). Higher variability was observed in GC than in LC, likely due to matrix interferences or  
268 compounds degradation. Indeed, the high variability was observed for compounds with low signal intensity  
269 and/or wide peak shape, which significantly influence the peak integration and thus the reported peak area.

270 Linearity, estimated as the correlation between the peak area and the concentration in spiked samples, was  
271 also satisfactory, with  $R^2$  values higher than 0.99 for all analysed compounds either in GC or LC, except for  
272  $\alpha$ -HBCDD, deltametrin and BTBPE for which  $R^2$  was greater than 0.97 (Table 1).

273 As shown in Table 1, no significant matrix effect was observed with regard to the LC-ESI(+/-)-HRMS  
274 detection, for which the purification strategy was satisfactory. Conversely, the GC-EI-HRMS detection was  
275 influenced by significant matrix effect, HCB, p-TBX, HBBz diagnostic signals being decreased between 40  
276 and 60%. Quizalofop-p-ethyl and deltametrin signals were increased 216 and 95%, respectively. These  
277 signal fluctuations could probably be reduced with more purified extracts to decrease matrix effect but it  
278 would lead to more selective sample preparation which is contrary to non-targeted approaches. New sample  
279 preparation approaches then appears necessary for NTS in order to combine an extended range of accessible  
280 biomarkers of exposure with a sufficiently clean extract for analysis. This requires the implementation of  
281 new preventive maintenance strategies such as guard-column and/or pre-filter and either in LC or GC. In LC  
282 a regular column clean-up after each sample batch is recommended, in the present case with  
283 isopropanol/acetone 1/1 (v/v), to insure system robustness and prevent peak resolution degradation by  
284 eliminating residues of matrix interferences potentially remaining in the system.

285 As shown in Table 1, 86%, 100% and 57% of compounds were well recovered using the LC-ESI(-), LC-  
286 ESI(+) and GC-EI methods respectively. Molecules with lower recovery, including 2,4-DCP, 2,4-DBP, p-  
287 TBX, chlorfenvinphos, PBDE 153, BTBPE, and anti-DP, informed on the method application range  
288 limitation.

289 Finally, all the test reference compounds passed the qualitative detection criterion at different concentration  
290 levels, except 2,4-DCP, which was not detected and 2,4-DBP and BTBPE which were only detected at the  
291 two highest concentration levels. A subset of 19 compounds also passed the linearity, ME and recovery  
292 criteria as summarised in Table 2. In order to better understand the root causes of these method limitations in  
293 terms of accessible markers, a statistical modelling approach was conducted to predict the expected recovery  
294 from the markers' physicochemical properties.

## 295 2.2 Multivariate regression model to predict recovery from molecular structures

296 A quantitative structure-recovery relationship (qsRecr) model was built using the molecular descriptors of  
297 the compounds regressed versus their recovery obtained with the analytical method developed in this study.  
298 The model's  $R^2$  and  $Q^2$  coefficients were equal to 0.77 and 0.50 (respectively) with a p-value below 0.05  
299 (0.0052), indicating a significant reliability. The model was able to predict with good accuracy the recovery  
300 observed for our different reference compounds, with a RMSE and a cross-validated RMSE (RMSEcv) of  
301 0.087 and 0.118, respectively. Predicted versus experimentally observed recovery for 30 test compounds  
302 used as training set (blue dots) in the OPLS model are reported in Figure 3. As shown, only three compounds  
303 (10%) are outside the RMSEcv limits. This means that the model is accurate in 90% of the cases, regardless  
304 of the structural variability of the 30 compounds in the training set.

305 The regression model built in this study can thus be used to predict the recovery of compounds non-available  
306 as chemical standards, using as input data the structural descriptors retrieved from PubChem. To corroborate  
307 this finding and evaluate over-fitting, the model was tested against a test-set of three compounds: atrazine,  
308 PFOS and BDCIPP. Their recovery was predicted using the model and is plotted in Figure 3 (red dots). As  
309 shown, the error of prediction for the three compounds was 0.106, 0.032 and 0.038, respectively. This result  
310 confirmed the accuracy of the qsRecr model to predict the compounds' recovery in the developed analytical  
311 method, with a relatively small error, based only on their chemical structures and associated physicochemical  
312 properties.

313 The results of the OPLS model can also identify which parameters most strongly influence the recovery of  
314 the compounds. Taking in account the limited number of compounds of the model (n=30), we report here a  
315 list of variables that may contribute to the predictive model (VIP > 1 and correlation coefficient higher than  
316 0.015) (Table S5). Four main observations can be drawn: (1) Recovery increases for unbranched cycle  
317 molecule when the mass, complexity or number of heavy atom increase. (2) The higher the molecular mass  
318 (>400-500 Da) or the log P (log P >5-6), the lower the recovery, except if the molecule contains OH group or  
319 oxygen double bond. (3) The higher the molecular mass and/or the log P of a molecule containing heavy  
320 atoms such as bromine, the lower the recovery. (4) Following the two last observations, the higher the degree  
321 of bromination, the lower the recovery except if the chemical structure contains OH group or oxygen double

322 bond. These observations indicate limitations of the method for high molecular mass and/or hydrophobic  
323 molecules, especially when containing bromine atoms, except if the molecule contains OH group or oxygen  
324 double bond. This is in agreement with the fact that Captiva EMR-Lipid sorbent selectivity is based on size  
325 exclusion and hydrophobic interaction, which could explain the low recoveries for PBDE 153, anti-DP and  
326 BTBPE. However, low recoveries were also observed for smaller and more polar molecules as HCB and p-  
327 TBX. Signal of these molecules also decreased by over 30% according to the matrix effect calculation. We  
328 hypothesised that the compounds could be well detected with the present analytical conditions in a clean  
329 system. However, the presence of matrix could impair their detection and this phenomenon is increased with  
330 high injector temperature and a long transfer line (more > 40 cm) which are parts of the system known for  
331 compound degradation. This highlights the complexity of non-targeted method development to combine a  
332 non-selective sample preparation with analytical system detecting a wide range of molecules. Thanks to the  
333 results of the qsRecr model, another NTS method with complementary performance can be developed, with a  
334 specific focus on GC detection.

335 The elaborated qsRecr model appears as an innovative approach to predict the recovery of given exposure  
336 markers with regard to a given sample preparation strategy based on their physicochemical properties. The  
337 continuous inclusion of further experimental results will contribute to increase the robustness of this model.  
338 Such model can also be used for interlaboratory comparisons with regard to the efficiency of various NTS  
339 methods.

## 340 2.3 Proof-of-concept: real sample analysis

### 341 2.3.1 *Blank sample investigation*

342 The analytical batch devoted to the proof-of-concept included a triplicate of procedural blank samples where  
343 human milk was replaced by pure water. Deltametrin and (Z)-dimetomorph were detected in blanks and  
344 matrix samples analysed by GC with the same order of magnitude for peak area. However, since the  
345 procedural contamination was not quantified the, signals detected in at least one blank sample was discarded  
346 and these two compounds were not considered for further interpretations. This observation highlights the  
347 importance of procedural blanks especially for non-targeted screening. It also leads to new compromise

348 when ubiquitous compounds are detected in procedural blanks and also in the original sample, especially if  
349 the blank is more contaminated than the sample because it does not perfectly mimic the matrix.

### 350 2.3.2 *Human milk sample investigation*

351 The number of clusters aligned in triplicate composite of milk analysed in LC-ESI(-) and LC-ESI(+) was  
352 reduced from 286 and 205 to 45 and 12 clusters after blank subtraction and manual investigation,  
353 respectively. Then, 17 and 5 clusters matched with the HBM4EU list. Among the 17 hits detected in LC-  
354 ESI(-), an isotopic cluster matched with 4-hydroxychlorothalonil, a metabolite of chlorothalonil (pesticide).  
355 Mass spectra (experimental and theoretical) matched at 91% (score obtained from HaloSeeker v1.0) with 0.4  
356 ppm of mass deviation. The pure analytical standard of 4-hydroxychlorothalonil was analysed in the same  
357 chromatographic conditions and compounds were eluted at the same retention time (difference lower than  
358 0.1 min). Both fragmentation mass spectra were generated at normalised collision energy of 60%. Resulting  
359 peak intensities and mass deviation were compared and matched perfectly (Figure 4). According to the  
360 confidence level proposed by Schymanski *et al.* [37], the 4-hydroxychlorothalonil was identified in human  
361 milk starting from a non-targeted approach at confidence level 1.

362 From the generated GC-HRMS data, HCB and p,p'-DDE were identified at confidence level 1, according to  
363 the Schymanski scale (Figure 5). A semi-quantification approach was possible because the method has been  
364 assessed with these chemicals. However, both compounds were detected at lower concentration than the  
365 calibration curve. Thus, HCB and p,p'-DDE were detected at concentrations lower than 6.25 ng mL<sup>-1</sup>. These  
366 two chemicals have already been detected in human milk in several studies [9,38,39], which validates the  
367 present method's capacity to detect environmental contaminants. However, this method faces some  
368 limitations regarding nature and concentration of accessible markers, especially in the identification step in  
369 GC-HRMS which requires advanced software addressing halogenated pattern issue and real database with  
370 experimental and/or *in-silico* fragmentation spectra.

## 371 **3 Conclusions**

372 Thanks to NTS approaches, we can gradually extend the knowledge of the human exposure and more  
373 broadly characterise the pressures exerted by environmental chemicals in the framework environment-food-



374 human health. In this study, a NTS workflow based on complementary LC- and GC-HRMS platforms was  
375 developed, assessed and applied to the analysis of a human milk composite sample. Method performance,  
376 including linearity, recovery, matrix effects and sensitivity was assessed for a reference mix of 30  
377 compounds. Although this group of molecules represents a limited range of the existing contaminants in the  
378 chemical universe, it allowed us to determine some methods limitations and efficiencies. The elaborated  
379 qsRecr model was able to predict with good accuracy the recovery of novel identified compounds based on  
380 their physicochemical properties. In general, the model indicated that this method may insufficiently detect  
381 bigger and less polar molecules without alcoholic or ketone groups. This model is also an innovative  
382 approach for documenting method limitations. It also illustrates the need of pluridisciplinary knowledge,  
383 including analytical chemistry, computational modelling and statistics, to properly develop NTS method.  
384 Beyond the complementarity and useful integration of both LC and GC platforms to cover a broad range of  
385 molecules, the current performance of NTS approaches still appears below those of more specific targeted  
386 methods especially for heavy and brominated compounds. The comprehensive analysis of human milk with  
387 LC and GC was able to detect and identify 4-hydroxychlorothalonil, p,p'-DDE and HCB with a single  
388 sample preparation, respectively. To conclude, our results have demonstrated that the developed analytical  
389 strategy is effective for the non-targeted monitoring of environmental chemical contaminants This approach  
390 will rapidly be able to generate internal chemical exposure and to contribute to the widening of knowledge of  
391 the human exposome.

#### 392 **4 Acknowledgements**

393 The authors thank the European Union's Horizon 2020 research and innovation programme HBM4EU under  
394 Grant Agreement No. 733032 for its financial support, as well as the Région Pays de la Loire, France, for its  
395 co-funding of M.P. Ph.D. grant. The authors thank Komodo Matta for language editing and proofreading.

#### 396 **5 References**

397 [1] C. P. Wild, Complementing the genome with an "Exposome": The outstanding challenge of  
398 environmental exposure measurement in molecular epidemiology. *Cancer Epidemiol. Biomarkers Prev.* 14  
399 (2005) 1847-1850. <https://doi.org/10.1158/1055-9965.EPI-05-0456>

- 400 [2] J. R. Sobus, Y. M. Tan, J. D. Pleil, L. S. Sheldon, A biomonitoring framework to support exposure and  
401 risk assessments. *Sci. Total Environ.* 409 (2011) 4875-4884. <https://doi.org/10.1016/j.scitotenv.2011.07.046>
- 402 [3] A. M. Calafat, Biomonitoring to assess exposures to mixtures of environmental chemicals, in: C. V.  
403 Rider, J. E. Simmons, *Chemical mixtures and combined chemical and nonchemical stressors: Exposure,*  
404 *Toxicity, Analysis, and Risk*, Springer International Publishing: Cham, 2018, pp. 15-36. Reference to the  
405 website: <https://www.springer.com/gp/book/9783319562322>
- 406 [4] M. M. Niedzwiecki, D. I. Walker, R. Vermeulen, M. Chadeau-Hyam, D. P. Jones, G. W. Miller, The  
407 Exposome: molecules to populations. *Annu. Rev. Pharmacol. Toxicol.* 59 (2019) 107-127.  
408 <https://doi.org/10.1146/annurev-pharmtox-010818-021315>
- 409 [5] M. Pruvost-Couvreur, B. Le Bizec, C. Béchaux, G. Rivière, Dietary risk assessment methodology: how to  
410 deal with changes through life. *Food Addit. Contam. Part A Chem. Anal. Control Expo. Risk Assess.* (2020)  
411 1-18. <https://doi.org/10.1080/19440049.2020.1727964>
- 412 [6] C. Ganzleben, J. P. Antignac, R. Barouki, A. Castaño, U. Fiddicke, J. Klánová, E. Leuret, N. Olea, D.  
413 Sarigiannis, G. R. Schoeters, Human biomonitoring as a tool to support chemicals regulation in the European  
414 Union. *Int. J. Hyg. Environ. Health*, 220 (2017) 94. <https://doi.org/10.1016/j.ijheh.2017.01.007>
- 415 [7] World Health Organization. Principles for evaluating health risks in children associated with exposure to  
416 chemicals, 2006. Reference to the website: <https://apps.who.int/iris/handle/10665/43604>
- 417 [8] E. Vizcaino, J. O. Grimalt, A. Fernández-Somoano, A. Tardon, Transport of persistent organic pollutants  
418 across the human placenta. *Environ. Int.*, 65 (2014) 107-115. <https://doi.org/10.1016/j.envint.2014.01.004>
- 419 [9] P. Fürst, Dioxins, polychlorinated biphenyls and other organohalogen compounds in human milk. Levels,  
420 correlations, trends and exposure through breastfeeding. *Mol. Nutr. Food Res.* 50 (2006) 922-933.  
421 <https://doi.org/10.1002/mnfr.200600008>
- 422 [10] C. S. Alcalá, L. J. Phillips, PCB concentrations in women based on breastfeeding history: NHANES  
423 2001–2004. *Environ. Res.* 154 (2017) 35-41. <https://doi.org/10.1016/j.envres.2016.12.015>

- 424 [11] A. Kyriklaki, M. Vafeiadi, M. Kampouri, K. Koutra, T. Roumeliotaki, G. Chalkiadaki, D. Anousaki, P.  
425 Rantakokko, H. Kiviranta, E. Fthenou, P. Bitsios, S. A. Kyrtopoulos, M. Kogevinas, L. Chatzi, Prenatal  
426 exposure to persistent organic pollutants in association with offspring neuropsychological development at  
427 4years of age: The Rhea mother-child cohort, Crete, Greece. *Environ. Int.* 97 (2016) 204-211.  
428 <https://doi.org/10.1016/j.envint.2016.09.012>
- 429 [12] S. A. Berghuis, K. N. J. A. van Braeckel, P. J. J. Sauer, A. F. Bos, Prenatal exposure to persistent  
430 organic pollutants and cognition and motor performance in adolescence. *Environ. Int.*, 121 (2018) 13-22.  
431 <https://doi.org/10.1016/j.envint.2018.08.030>
- 432 [13]W. H. Oddy, Breastfeeding protects against illness and infection in infants and children: a review of the  
433 evidence. *Breastfeeding Rev.* 9 (2001) 8-11.
- 434 [14] L. L. Needham, R. Y. Wang, Analytic considerations for measuring environmental chemicals in breast  
435 milk. *Environ. Health Perspect.* 110 (2002) A317. <https://doi.org/10.1289/ehp.021100317>
- 436 [15] S. A. Lee, Q. Dai, W. Zheng, Y. T. Gao, A. Blair, J. D. Tessari, B. Tian Ji, X. O. Shu, Association of  
437 serum concentration of organochlorine pesticides with dietary intake and other lifestyle factors among urban  
438 Chinese women. *Environ. Int.* 33 (2007) 157-163. <https://doi.org/10.1016/j.envint.2006.08.010>
- 439 [16] S. N. Sinha, V. R. Banda, Correlation of pesticide exposure from dietary intake and bio-monitoring: The  
440 different sex and socio-economic study of children. *Ecotoxicol. Environ. Saf.* 162 (2018) 170-177.  
441 <https://doi.org/10.1016/j.ecoenv.2018.06.086>
- 442 [17] J. M. Weiss, E. Simon, G. J. Stroomborg, R. de Boer, J. de Boer, S. C. van der Linden, P; E. G.  
443 Leonards, M. H. Lamoree, Identification strategy for unknown pollutants using high-resolution mass  
444 spectrometry: Androgen-disrupting compounds identified through effect-directed analysis. *Anal. Bioanal.*  
445 *Chem.* 400 (2011) 3141-3149. <https://doi.org/10.1007/s00216-011-4939-x>
- 446 [18] E. Simon, M. van Velzen, S. H. Brandsma, E. Lie, K. Løken, J. de Boer, J. Bytingsvik, B. M. Jenssen, J.  
447 Aars, T. Hamers, M. H. Lamoree, Effect-fired analysis to explore the polar bear exposome: identification

448 of thyroid hormone disrupting compounds in plasma. *Environ. Sci. Technol.* 47 (2013) 8902-8912.  
449 <https://doi.org/10.1021/es401696u>

450 [19] K. M. Blum, P. L. Andersson, G. Renman, L. Ahrens, M. Gros, K. Wiberg, P. Haglund, Non-target  
451 screening and prioritization of potentially persistent, bioaccumulating and toxic domestic wastewater  
452 contaminants and their removal in on-site and large-scale sewage treatment plants. *Sci. Total Environ.* 575  
453 (2017) 265-275. <https://doi.org/10.1016/j.scitotenv.2016.09.135>

454 [20] J. Hollender, E. L. Schymanski, H. P. Singer, P. L. Ferguson, Nontarget screening with high resolution  
455 mass spectrometry in the environment: ready to go? *Environ. Sci. Technol.* 51 (2017) 11505-11512.  
456 <https://doi.org/10.1021/acs.est.7b02184>

457 [21] Y. Fu, C. Zhao, X. Lu, G. Xu, Nontargeted screening of chemical contaminants and illegal additives in  
458 food based on liquid chromatography–high resolution mass spectrometry. *TrAC Trends in Anal. Chem.* 96  
459 (2017) 89-98. <https://doi.org/10.1016/j.trac.2017.07.014>

460 [22] E. Omer, R. Cariou, G. Remaud, Y. Guitton, H. Germon, P. Hill, G. Dervilly-Pinel, B. Le Bizec,  
461 Elucidation of non-intentionally added substances migrating from polyester-polyurethane lacquers using  
462 automated LC-HRMS data processing. *Anal. Bioanal. Chem.* 410 (2018), 5391-5403.  
463 <https://doi.org/10.1007/s00216-018-0968-z>

464 [23] M. Pourchet, L. Debrauwer, J. Klanova, E. J. Price, A. Covaci, N. Caballero-Casero, H. Oberacher, M.  
465 H. Lamoree, A. Damont, F. Fenaille, J. Vlaanderen, J. Meijer, M. Krauss, D. Sarigiannis, R. Barouki, B. Le  
466 Bizec, J. P. Antignac, Suspect and non-targeted screening of chemicals of emerging concern for human  
467 biomonitoring, environmental health studies and support to risk assessment: From promises to challenges  
468 and harmonisation issues. *Environ. Int.* 139 (2020) 105545. <https://doi.org/10.1016/j.envint.2020.105545>

469 [24] R. Cariou, E. Omer, A. Léon, G. Dervilly-Pinel, B. Le Bizec, Screening halogenated environmental  
470 contaminants in biota based on isotopic pattern and mass defect provided by high resolution mass  
471 spectrometry profiling. *Anal. Chim. Acta* 936 (2016) 130-138. <https://doi.org/10.1016/j.aca.2016.06.053>

472 [25] A. Léon, R. Cariou, S. Hutinet, J. Hurel, Y. Guitton, C. L. Tixier, C. Munsch, J. P. Antignac, G.  
473 Dervilly-Pinel, B. Le Bizec, HaloSeeker 1.0: A user-friendly software to highlight halogenated chemicals in  
474 nontargeted high-resolution mass spectrometry data sets. *Anal. Chem.* 91 (2019) 3500-3507.  
475 <https://doi.org/10.1021/acs.analchem.8b05103>

476 [26] R. Kaliszan, Quantitative structure-retention relationships. *Anal. Chem.* 64 (1992) 11.  
477 <https://doi.org/10.1021/ac00035a001>

478 [27] M. C. Alexandre-Gouabau, T. Moyon, A. David-Sochard, F. Fenaille, S. Cholet, A. L. Royer, Y.  
479 Guitton, H. Billard, D. Darmaun, J. C. Roze, C. Y. Boquien, Comprehensive preterm breast milk metabotype  
480 associated with optimal infant early growth pattern. *Nutrients* 11 (2019) 528.  
481 <https://doi.org/10.3390/nu11030528>

482 [28] G. Cano-Sancho, M. C. Alexandre-Gouabau, T. Moyon, T. A. L. Royer, Y. Guitton, H. Billard, D.  
483 Darmaun, J. C. Roze, C. Y. Boquien, B. Le Bizec, J. P. Antignac, Simultaneous exploration of nutrients and  
484 pollutants in human milk and their impact on preterm infant growth: An integrative cross-platform approach.  
485 *Environ. Res.* 182 (2020) 109018. <https://doi.org/10.1016/j.envres.2019.109018>

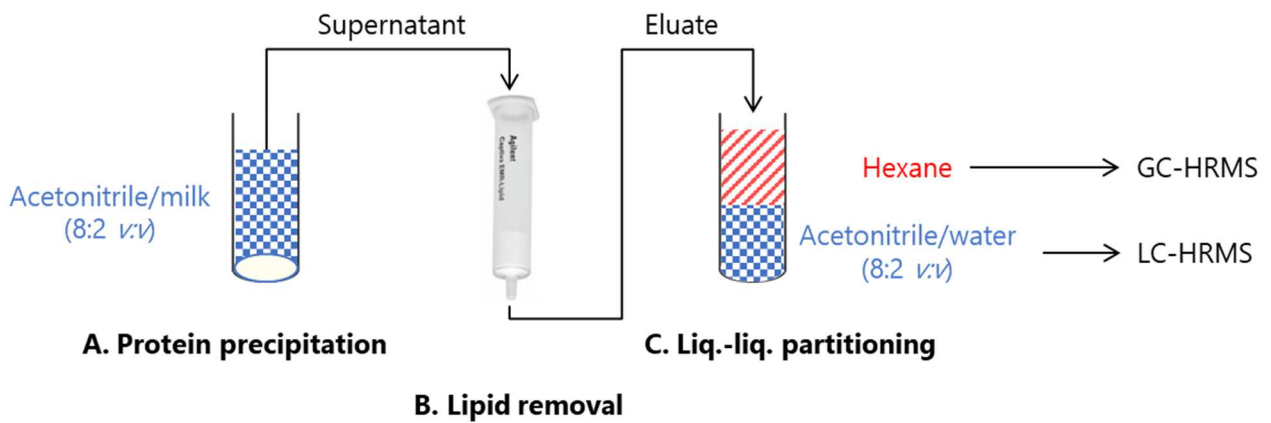
486 [29] L. Han, J. Matarrita, Y. Sapozhnikova, S. J. Lehotay, Evaluation of a recent product to remove lipids  
487 and other matrix co-extractives in the analysis of pesticide residues and environmental contaminants in  
488 foods. *J. Chrom. A.* 1449 (2016) 17-29. <https://doi.org/10.1016/j.chroma.2016.04.052>

489 [30] L. Zhao, T. Szakas, M. Churley, D. Lucas, Multi-class multi-residue analysis of pesticides in edible oils  
490 by gas chromatography-tandem mass spectrometry using liquid-liquid extraction and enhanced matrix  
491 removal lipid cartridge cleanup. *J. Chrom. A.* 1584 (2019) 1-12.  
492 <https://doi.org/10.1016/j.chroma.2018.11.022>

493 [31] B. Arce-López, E. Lizarraga, M. Flores-Flores, A. Irigoyen, E. González-Peñas, Development and  
494 validation of a methodology based on Captiva EMR-lipid clean-up and LC-MS/MS analysis for the  
495 simultaneous determination of mycotoxins in human plasma. *Talanta* 206 (2020) 120193.  
496 <https://doi.org/10.1016/j.talanta.2019.120193>

- 497 [32] D. Stettin, R. X. Poulin, G. Pohnert, Metabolomics benefits from Orbitrap GC–MS—comparison of  
498 low- and high-resolution GC–MS. *Metabolites* 10 (2020) 143. <https://doi.org/10.3390/metabo10040143>
- 499 [33] N. W. Kwiecien, D. J. Bailey, M. J. P. Rush, J. S. Cole, A. Ulbrich, A. S. Hebert, M. S. Westphall, J. J.  
500 Coon, High-resolution filtering for improved small molecule identification via GC/MS. *Anal. Chem.* 87  
501 (2015) 8328-8335. <https://doi.org/10.1021/acs.analchem.5b01503>
- 502 [34] S. Wold, N. Kettaneh-Wold, B. Skagerberg, Nonlinear PLS modelling. *Chemometr. Intel. Lab. Syst.* 7  
503 (1989) 53-65. [https://doi.org/10.1016/0169-7439\(89\)80111-X](https://doi.org/10.1016/0169-7439(89)80111-X)
- 504 [35] L. Sleno, The use of mass defect in modern mass spectrometry. *J. Mass Spectrom.* 47 (2012), 226-236.  
505 <https://doi.org/10.1002/jms.2953>
- 506 [36] L. Eriksson, J. Trygg, S. Wold, CV-ANOVA for significance testing of PLS and OPLS® models. *J.*  
507 *Chemometrics*, 22 (2008) 594-600. <https://doi.org/10.1002/cem.1187>
- 508 [37] E. L. Schymanski, H. P. Singer, J. Slobodnik, I. M. Ipolyi, P. Oswald, M. Krauss, T. Schulze, P.  
509 Haglund, T. Letzel, S. Grosse, N. S. Thomaidis, A. Bletsou, C. Zwiener, M. Ibáñez, T. Portolés, R. de Boer,  
510 M. J. Reid, M. Onghena, U. Kunkel, W. Schulz, A. Guillon, N. Noyon, G. Leroy, P. Bados, S. Bogialli, D.  
511 Stipaničev, P. Rostkowski, J. Hollender, Non-target screening with high-resolution mass spectrometry:  
512 critical review using a collaborative trial on water analysis. *Anal. Bioanal. Chem.* 407 (2015) 6237-6255.  
513 <https://doi.org/10.1007/s00216-015-8681-7>
- 514 [38] C. Campoy, M. Jiménez, M. F. Olea-Serrano, M. Moreno Frias, F. Cañabate, N. Olea, R. Bayés, J. A.  
515 Molina-Font, Analysis of organochlorine pesticides in human milk: preliminary results. *Early Hum. Dev.* 65  
516 (2001) S183-S190. [https://doi.org/10.1016/s0378-3782\(01\)00221-3](https://doi.org/10.1016/s0378-3782(01)00221-3)
- 517 [39] R. Aerts, I. Van Overmeire, A. Colles, M. Andjelković, G. Malarvannan, G. Poma, E. Den Hond, E. van  
518 de Mierop, M. C. Dewolf, F. Charlet, A. van Nieuwenhuysse, J. van Loco, A. Covaci, Determinants of  
519 persistent organic pollutant (POP) concentrations in human breast milk of a cross-sectional sample of  
520 primiparous mothers in Belgium. *Environ. Int.* 131 (2019) 104979.  
521 <https://doi.org/10.1016/j.envint.2019.104979>

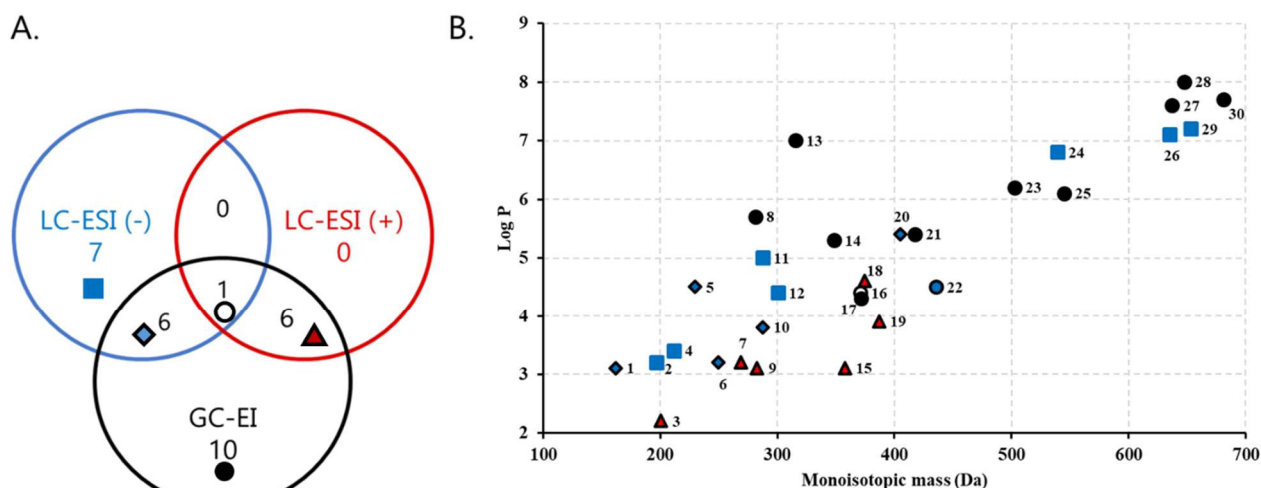
522



523

524 *Figure 1: Sample preparation protocol based on A. milk protein precipitation with acetonitrile, B. supernatant (water*  
525 *from the milk and acetonitrile) is loaded on the Captiva EMR-Lipid® cartridge to remove lipids. C. eluate*  
526 *(water/acetonitrile) is partitioning with 2x2 mL of hexane.*

527



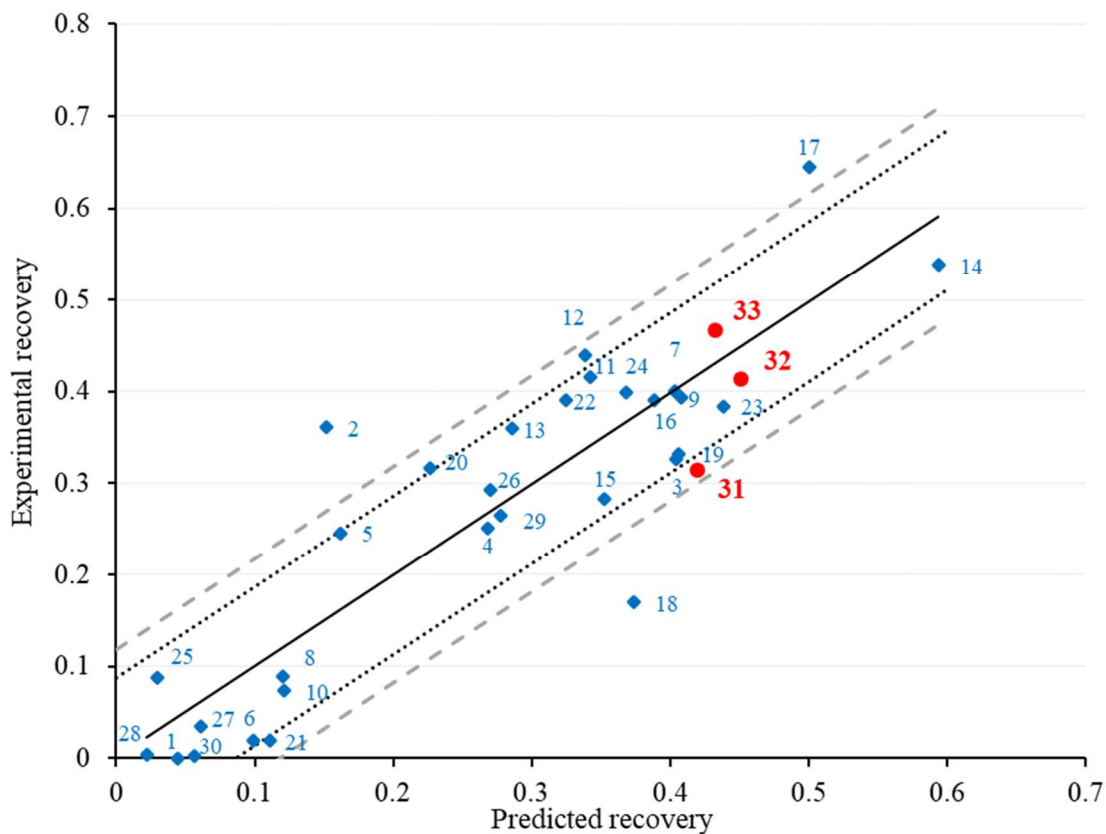
1 – 2,4-DCP; 2 – TCPy; 3 – Simazine; 4 – Fenvalerate free acid; 5 – 2,3,4,5-tetraCP; 6 – 2,4-DBP; 7 – Acetochlor; 8 – HCB; 9 – Metolachlor; 10 –  $\beta$ -HCH; 11 – Triclosan; 12 – Fenhexamid; 13 – p,p'-DDE; 14 – Chlorpyrifos; 15 – Chlorfenvinphos; 16 – Tetraconazole; 17 – Quinalofop-p-ethyl; 18 – Prochloraz; 19 – (Z)-Dimethomorph; 20 – 2,3,4,5-tetraBP; 21 – p-TBX; 22 – Fipronil; 23 – Deltamethrin; 24 – TBBPA; 25 – HBBz; 26 –  $\alpha$ -HBCDD; 27 – PBDE 153; 28 – a-DP; 29 – 6-OH-BDE 137; 30 – BTBPE

529

530 *Figure 2: Venn diagram of LC-ESI(+/-) and GC-EI complementarity to detect a wide range of molecules (30 test*  
 531 *reference compounds listed in Table 1 in SI) (A) with different physicochemical properties as monoisotopic mass and*  
 532 *polarity (B). Squarre and circle are compounds detected in LC-ESI(-) and GC-EI. Rhombus and triangle are*  
 533 *compounds detected in both LC-ESI(-)/GC-EI and LC-ESI(+)/GC-EI. The empty circle (no 16) is the compound detected*  
 534 *by three modes.*

535

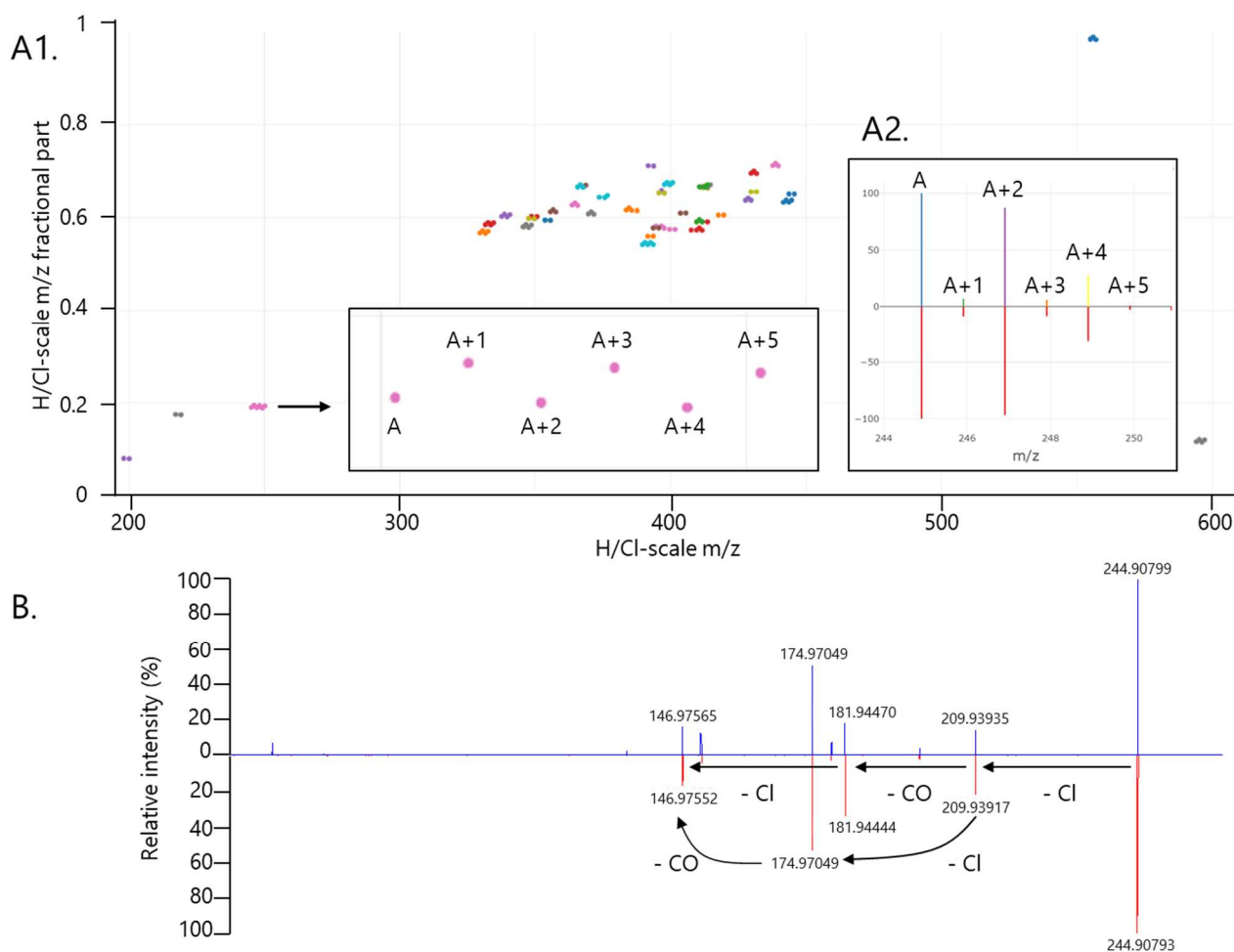




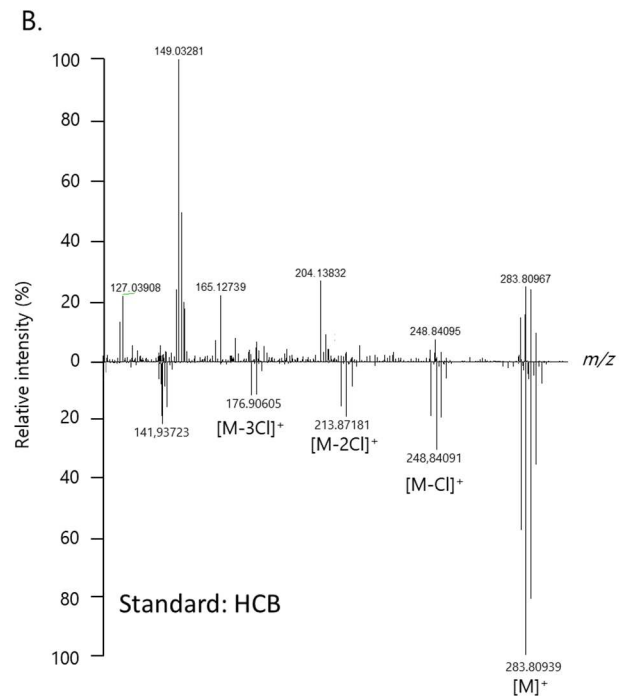
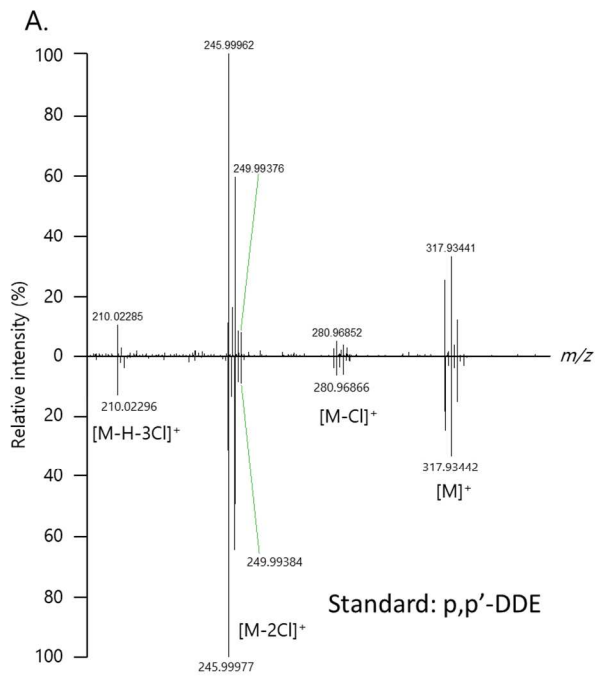
536

537 *Figure 3: Results of the OPLS model built to predict recovery from a set of physicochemical properties of the*  
 538 *considered biomarkers of exposure. Compound numbers used to build the model are referred in figure 1 and 31 –*  
 539 *atrazine; 32 – BDCIPP; 33 – PFOS were test compounds to assess model accuracy. Equation of the linear regression*  
 540 *curve is  $Y = 0.9942x + 0.0006$ ;  $r^2 = 0.77$ ;  $RMSE_{min} = 0.087$  and  $RMSE_{max} = 0.118$  are represented as upper and lower*  
 541 *limits in black dotted line and grey dash line, respectively.*

542



545 *Figure 4: Identification of an unknown halogenated cluster detected in unspiked human milk sample by LC-ESI(-)-*  
 546 *HRMS. (A1) Mass defect plot generated by HaloSeeker v1.0. (A2) Experimental (top) and theoretical (bottom) mass*  
 547 *spectra of the compound  $C_8HCl_3N_2O$  (score of 91% and mass deviation 0.4 ppm). B. fragmentation mass spectrum of the*  
 548 *unknown cluster at NCE 60% (B top) compared with the mass spectrum fragmentation of the pure analytical standard*  
 549 *of hydroxy-chlorothalonil (B bottom) Both compounds were eluted at 4.63 ( $\pm 0.1$ ) min. Peak list with mass deviation in*  
 550 *ppm in brackets: 244.90799 (-0.7); 209.93935 (0.2); 181.94470 (1.6); 174.97049 (0.2) and 146.97565 (0.7).*



552

553 *Figure 5: Fragmentation mass spectra obtained by GC-EI-HRMS of p,p'-DDE (A) and HCB (B) in human milk (top)*  
 554 *and the analytical standard (bottom). Peak list with mass deviation in ppm in brackets for p,p'-DDE: 317.93441 (0.03);*  
 555 *280.96852 (0.5); 245.99962 (0.3); 210.02285 (0.5) and for HCB: 283.80967 (-1.0); 248.84095 (-0.1).*

556

557

Table 1: Sample preparation linearity and matrix effect (ME) in GC-EI and LC-ESI(+/-). The mean of ME observed for sample in triplicate is reported.

Compounds name	GC-EI		LC-ESI(-)		LC-ESI(+)		Recovery			
	r <sup>2</sup>	ME	r <sup>2</sup>	ME	r <sup>2</sup>	ME	Level 1	Level 2	Level 3	Level 4
<b>2,4-DCP</b>	-	-	ND	-16%	-	-	ND	ND	ND	ND
<b>2,4-DBP</b>	-	-	0.997	+23%	-	-	ND	ND	2%	3%
<b>2,3,4,6-tetra-CP</b>	-	-	0.993	+6%	-	-	20%	13%	24%	24%
<b>HCB</b>	0.994	-58%	-	-	-	-	6%	2%	5%	7%
<b>Simazine</b>	-	-	-	-	0.999	-5%	33%	28%	31%	31%
<b>β-HCH</b>	0.998	-23%	-	-	-	-	5%	3%	3%	6%
<b>Acetochlor</b>	0.997	-23%	-	-	0.997	0%	28%	22%	39%	33%
<b>2,3,4,6-tetra-BP</b>	-	-	0.996	0%	-	-	27%	23%	31%	32%
<b>Metolachlor</b>	0.997	-25%	-	-	0.995	-2%	33%	24%	39%	32%
<b>Chlorpyrifos</b>	0.998	-10%	-	-	-	-	39%	20%	20%	38%
<b>Tetraconazole</b>	-	-	0.995	+2%	1.000	+3%	37%	27%	38%	35%
<b>Fipronil</b>	-	-	0.997	-3%	-	-	36%	29%	38%	37%
<b>p-TBX</b>	0.990	-51%	-	-	-	-	2%	0%	1%	1%
<b>Chlorfenvinphos</b>	-	-	-	-	0.999	-2%	29%	21%	28%	28%
<b>p,p'-DDE</b>	0.989	-12%	-	-	-	-	23%	11%	17%	24%
<b>HBBz</b>	0.991	-40%	-	-	-	-	2%	2%	4%	5%
<b>Prochloraz</b>	-	-	-	-	0.999	0%	17%	12%	16%	17%
<b>Quizalofop-p-ethyl</b>	0.998	+216%	-	-	-	-	31%	26%	10%	16%
<b>Deltametrin</b>	0.972	+95%	-	-	-	-	20%	22%	5%	12%
<b>PBDE 153</b>	0.996	-29%	-	-	-	-	1%	0%	1%	2%
<b>(Z)-Dimetomorph</b>	-	-	-	-	0.999	+2%	33%	24%	32%	33%
<b>BTBPE</b>	0.978	+19%	-	-	-	-	0%	0%	0%	0%
<b>anti-DP</b>	0.990	-29%	-	-	-	-	0%	0%	0%	0%
<b>OH-BDE 137</b>	-	-	0.998	-6%	-	-	20%	19%	26%	28%
<b>TBBPA</b>	-	-	0.997	+25%	-	-	29%	28%	41%	43%
<b>α-HBCDD</b>	-	-	0.973	+21%	-	-	12%	12%	28%	23%
<b>TCPy</b>	-	-	0.992	+11%	-	-	34%	27%	35%	32%
<b>Fenvalerate free acid</b>	-	-	0.992	+6%	-	-	11%	9%	24%	21%
<b>Fenhexamid</b>	-	-	0.996	+3%	-	-	40%	36%	44%	42%
<b>Triclosan</b>	-	-	0.994	+26%	-	-	40%	28%	40%	37%

ND: Not detected

560 Table 2: Subset group of reference test compounds with validated criteria. The mean of ME observed for sample in  
 561 triplicate is reported.

	Compounds name	LOD	Linearity	ME
<b>GC-EI</b>	<b>Acetochlor</b>	0.005	0.997	-23%
	<b>Metolachlor</b>	0.01	0.997	-25%
	<b>Chlorpyrifos</b>	0.005	0.998	-10%
	<b>p,p'-DDE</b>	0.001	0.989	-12%
<b>LC-ESI(-)</b>	<b>2,3,4,5-tetra-CP</b>	0.001	0.993	+6%
	<b>2,3,4,5-tetra-BP</b>	0.005	0.996	0%
	<b>Tetraconazole</b>	0.001	0.995	+2%
	<b>Fipronil</b>	0.001	0.997	-3%
	<b>OH-BDE 137</b>	0.001	0.998	-6%
	<b>TBBPA</b>	0.005	0.997	+25%
	<b>α-HBCDD</b>	0.01	0.973	+21%
	<b>TCPy</b>	0.01	0.992	+11%
	<b>Fenvalerate free acid</b>	0.1	0.992	+6%
	<b>Fenhexamid</b>	0.005	0.996	+3%
	<b>Triclosan</b>	0.001	0.994	+26%
	<b>LC-ESI(+)</b>	<b>Simazine</b>	0.001	0.999
<b>Acetochlor</b>		0.01	0.997	0%
<b>Metolachlor</b>		0.001	0.995	-2%
<b>Tetraconazole</b>		0.01	1.000	+3%
<b>Chlorfenvinphos</b>		0.01	0.999	-2%
<b>Prochloraz</b>		0.001	0.999	0%
<b>(Z)-Dimetomorph</b>		0.005	0.999	+2%

LOD: Limit of detection in ng μL<sup>-1</sup>

ME: Matrix effect

562

563

564

565

566

567

568

569

570

571

572

573

574

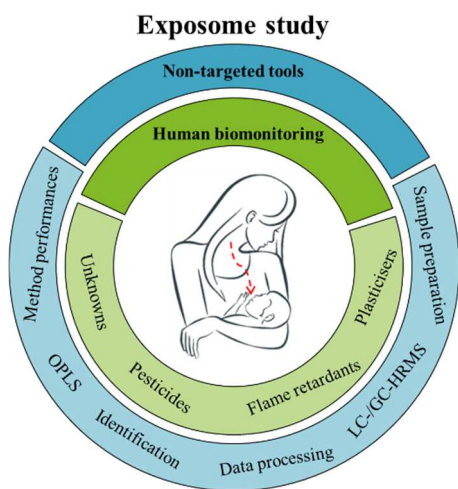
575

576

577

578

579 **Graphical abstract**



580