Title

Machine Learning-Assisted Identification and Quantification of Hydroxylated Metabolites of Polychlorinated Biphenyls in Animal Samples

Journal

Environmental Science and Technology, 56(18)

ISSN

0013-936X

Authors

Zhang, Chun-Yun
Li, Xueshu
Stietz, Kimberly P Keil
et al.

Publication Date

2022-09-20

DOI

10.1021/acs.est.2c02027

MACHINE LEARNING-ASSISTED IDENTIFICATION AND QUANTIFICATION OF HYDROXYLATED METABOLITES OF POLYCHLORINATED BIPHENYLS IN ANIMAL SAMPLES

Chun-Yun Zhang[1], Xueshu Li[1], Kimberly P. Keil Stietz[2], Sunjay Sethi[2], Weizhu Yang[3], Rachel F. Marek,[4] Xinxin Ding[3], Pamela J. Lein[2], Keri C. Hornbuckle[4], Hans-Joachim Lehmler[1,*]

[1]Department of Occupational and Environmental Health, The University of Iowa, Iowa City, Iowa 52242, United States. [2]Department of Molecular Biosciences, School of Veterinary Medicine, University of California Davis, Davis, California 95616, United States. [3]Department of Pharmacology and Toxicology, College of Pharmacy, University of Arizona, Tucson, Arizona 85721, United States. [4]Department of Civil and Environmental Engineering and IIHR Hydroscience and Engineering, The University of Iowa, Iowa City, IA 52242, United States.

Corresponding Author:

Dr. Hans-Joachim Lehmler

The University of Iowa

Department of Occupational and Environmental Health

University of Iowa Research Park, #164 MTF

Iowa City, IA 52242-5000

Phone: (319) 335-4981

Fax: (319) 335-4290

e-mail: hans-joachim-lehmler@uiowa.edu

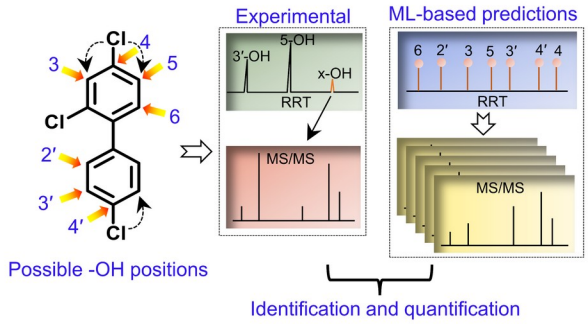
ABSTRACT

Laboratory studies of the disposition and toxicity of hydroxylated polychlorinated biphenyl (OH-PCB) metabolites are challenging because authentic analytical standards for most unknown OH-PCBs are not available. To assist with the characterization of these OH-PCBs (as methylated derivatives), we developed machine learning-based models with multiple linear regression (MLR) or random forest regression (RFR) to predict the relative retention times (RRT) and MS/MS responses of methoxylated (MeO-) PCBs on a gas chromatograph-tandem mass spectrometry (GC-MS/MS) system. The final MLR model estimated the retention times of MeO-PCBs with a mean absolute error of 0.55 min (n = 121). The similarity coefficients cos θ between the predicted (by RFR model) and experimental MS/MS data of MeO-PCBs were > 0.95 for 92% observations (n = 96). The levels of MeO-PCBs quantified with the predicted MS/MS response factors approximated the experimental values within a 2-fold difference for 85% of observations and 3-fold differences for all observations (n = 89). Subsequently, these model predictions were used to assist with the identification of OH-PCB 95 or OH-PCB 28 metabolites in mouse feces or liver by suggesting candidate ranking information for identifying the metabolite isomers. Thus, predicted retention and MS/MS response data can assist in identifying unknown OH-PCBs.

KEYWORDS: OH-PCBs; GC-MS/MS method; Model Prediction; Relative retention times; Relative response factor

SYNOPSIS: Machine learning-based models were used to identify and quantify toxicologically relevant hydroxylated PCB metabolites in biological samples.

45 TOC art


Cl
4
3
5
Cl
6
2′
3′
4′
Cl
Possible -OH positions
Experimental
5-OH
3′-OH
x-OH
RRT
MS/MS
ML-based predictions
6 2′ 3 5 3′ 4′ 4
RRT
MS/MS
Identification and quantification

46

INTRODUCTION

PCBs are a class of environmental pollutants that can be transformed into hydroxylated PCBs (OH-PCBs) by reaction with hydroxyl radicals in the environment[1, 2] or via oxidation by cytochrome P450 enzymes in organisms.[3] OH-PCBs are also present in technical PCB mixtures.[4] A total of 837 mono-hydroxylated PCBs (mono-OH-PCBs) and thousands of di-hydroxylated PCBs (di-OH-PCBs) can be formed from the 209 possible PCB congeners.[3] The parent PCBs are still present in the environment, human diet, and humans[5-9] and can be found in consumer products, such as paints and silicon rubber.[10-13] Therefore, it is not surprising that many OH-PCB congeners have been detected in environmental or biological media.[4, 14-16] OH-PCBs are potentially more toxic than the corresponding parent PCBs.[3] For example, OH-PCBs can interact with nuclear transcription factors, such as the aryl hydrocarbon receptor, constitutive androstane receptor, and pregnane X receptor.[17, 18] They are endocrine-disrupting chemicals that, for example, inhibit estrogen sulfotransferase and bind to transthyretin.[18-22] Di-OH-PCBs are oxidation products of mono-OH-PCBs, with PCB catechols being central PCB metabolites in mammals.[23-25] Di-OH-PCB metabolites can be transformed into PCB quinones, reactive PCB metabolites that cause oxidative stress or covalently bind to DNA and other cellular targets.[26-29] Some PCB catechols are tumor initiators in the liver.[30, 31]

Despite the well-documented toxicity of OH-PCBs, their presence in environmental samples, wildlife, laboratory animals, and humans has not been fully characterized, partly because of the lack of authentic analytical standards. OH-PCBs are typically analyzed as methylated derivatives (MeO-PCBs) with gas chromatographic (GC) methods.[23, 32, 33] GC can also be used to identify and quantify other PCB metabolites, such as PCB sulfates, as MeO-PCBs after deconjugation and

derivatization.[34] GC coupled with tandem mass spectrometry (GC-MS/MS) is a useful method to quantify the MeO-PCBs because of its good separation, high selectivity, and low detection limits for this class of compounds.[4, 14, 15] However, only a small number of the 837 possible OH-PCB congeners, either as hydroxylated or methoxylated derivatives,[35] are commercially available. The lack of authentic analytical standards represents a challenge for environmental, human biomonitoring, metabolism, and toxicity studies.[25, 35, 36] For example, unknown OH-PCB are frequently detected in environmental and biological samples.[36-43] Computational approaches can facilitate the identification and quantification of OH-PCBs in environmental and biological samples. However, no method is currently available for identifying and quantifying these metabolites in any matrix.

Computational models trained with experimental observations represent an alternative approach for the nontarget analysis of diverse groups of chemicals. For example, models have been developed to predict the retention times and response factors of PCBs,[44, 45] polybrominated diphenyl ether,[46] and human endogenous metabolites.[47] In silico predictions can simulate the MS/MS spectra of chemicals to support the identification of unknown compounds.[48] Previously, unknown OH-PCBs were quantified in abiotic samples using the average response factor for the OH-PCB homolog group.[43] We have previously shown that mono-OH-PCBs without authentic analytical standards can be identified by homolog group and quantified in PCB-contaminated sediment using a semi-nontargeted approach. However, because our method could not identify the substitution patterns, and could not identify di-hydroxyl PCBs, it was of limited use for interpreting the metabolic products of PCB exposure in laboratory animals.[35]

In this study, we used 124 analytical mono/di-MeO-PCB standards to develop multiple linear regression (MLR) or random forest regression (RFR) models that predict the retention times and MS/MS response data of MeO-PCBs on a GC-MS/MS system. The predicted GC-MS/MS data were used to identify and quantify OH-PCB metabolites in samples from animal studies with toxicologically relevant PCBs.

EXPERIMENTAL SECTION

Laboratory methods. This study used machine learning-based approaches to identify and quantify the OH-PCBs detected in biological samples from PCB disposition and toxicity studies. The biological samples investigated include a feces sample from a PCB disposition study with mice acutely exposed to an individual PCB congener (PCB 95) and a liver sample from a PCB disposition study with mice sub-chronically exposed to a human-relevant PCB mixture. Briefly, adult mice were exposed to PCB 95 (1.0 mg/kg), a neurotoxic PCB,[49-52] in stripped corn oil or corn oil alone. Feces from dissected distal colon and rectum were collected 24 h after PCB 95 exposure for analysis. The liver sample was collected as part of a larger study assessing the effects of developmental exposure to a PCB mixture on multiple developmental outcomes.[53-55] The biological samples were extracted following a published procedure[41, 56, 57] and analyzed by GC-MS/MS. For details regarding the animal studies, the extraction, and GC-MS/MS analysis, see the Supporting Information.

Experimental determination of RRTs and MS/MS profiles. Because of the high chromatographic resolution, OH-PCBs are typically extracted from biological or environmental matrices, derivatized to MeO-PCBs, and analyzed by GC-MS/MS.[4, 14, 15, 58] We measured the RRTs and MS/MS profiles [expressed as the relative intensities of five multiple reaction

monitering (MRM) transitions] of two MeO-PCB standard solutions (Solution 1 containing 72 MeO-PCBs and Solution 2 containing 52 MeO-PCBs; see Supporting Information for additional information) using an Agilent 7890B gas chromatograph equipped with an SPB-Octyl capillary column (30 m length, 250 μm inner diameter, 0.25 μm film thickness; Supelco, Bellefonte, PA, USA), an Agilent 7000D Triple Quad and an Agilent 7693 sampler. These data were used as dependent variables for the model development. For additional details, see the Supporting Information.

Model development. The two-fold goal of the model is to predict the identity and calculate the concentration of mono- and di-hydroxy PCBs in laboratory samples. We used MLR and RFR machine learning-based algorithms to develop models for identifying and quantifying OH-PCBs. These models used experimental RRT and RRF data (the components of MS/MS profiles) as dependent variables and molecular descriptors (MDs) as predictors. For the generation of chemoinformatics and substitution pattern-based MDs of the 124 MeO-PCBs (Table S1), see the Supporting Information. All data analyses were performed in R (version 3.6.3).

Preliminary data inspection. Since the MLR, but not the RFR models, assume normal data distribution and homogeneity of data variance,[60] a preliminary data inspection was performed on all datasets used to predict the RRTs and RRFs of MeO-PCBs with the MLR model. Inspection of diagnostic plots [i.e., normal probability plots (Q-Q plots) and residual vs. fitted value plots] for the RRT predictions suggested that the assumptions of data normality and variance homogeneity were supported by the majority of the 112 observations in the training datasets (Fig. S1).

The training datasets used for predicting RRFs revealed non-linear relationships. Therefore, the measured RRFs were log-transformed to obtain normally distributed data and account for non-linear relationships. Potential outlier observations were removed by Cook's distance (CD) with the following cut-off: CD < 10-fold of averaged CD (assuming outliers have CDs substantially larger than the averaged CD by over an order of magnitude). As a result, 109 and 88 observations remained in the training datasets used to develop models to predict RRTs and RRFs. Coeluting MeO-PCBs in the training dataset were removed for the prediction of RRFs.

MLR model development. We used a repeated 10-fold cross-validation strategy[61, 62] to train and internally validate the MLR models used to predict the RRTs or RRFs of MeO-PCBs. First, MLR modeling underwent a predictor selection step to minimize the number of predictors and enhance model stability without sacrificing model performance. This step was performed with the *stepAIC* function in the *MASS* package (https://cran.r-project.org/web/packages/MASS/index.html). Next, predictors were optimized stepwise with the Akaike Information Criteria (AIC) for variable selection. Based on this optimization step, ten out of 105 MDs were selected to predict RRTs (Table S2), and sixteen to sixty-six out of 105 MDs were used to predict the RRFs of the five MS transitions.

The observations from each dataset were randomly divided into ten groups. Nine groups were used as the training dataset, and the remaining dataset was used for internal testing. The model training and testing were performed ten times to ensure that each group was used once as the testing dataset. The data grouping, model training, and internal testing were repeated five times to avoid biases in the initial random grouping of the datasets. Finally, MLR models with predictor coefficients and their deviations at the least root mean square error (RMSE) were

generated to predict RRTs or RRFs. The MLR models were evaluated by R^2 (RSQ), mean absolute error (MAE), and RMSE between the predicted and measured value and the prediction interval at the 95 % confidence level.

RFR model development. Initially, RFR models were constructed to predict RRTs or RRFs with all MDs as independent variables and experimental RRTs or RRFs as dependent variables using the R package *randomForest*. Approximately two-thirds of the MeO-PCBs were randomly selected as the internal training dataset, and the rest were used as the internal testing dataset. An importance value was assigned to each MD to evaluate its contribution to the prediction model. The model construction was repeated 100 times with randomly selected datasets to identify the top six ranked MDs for each iteration. The MDs that appeared > 50 times in these RFR models were chosen for further predictions (Table S3).

Subsequently, the parameters in the random forest algorithms, *ntree* (i.e., number of trees to grow) and *mtry* (i.e., number of variables randomly sampled as candidates at each split), were optimized from 100 to 1000 with a step size of 100 for *ntree* and from one variable to the total number of variables for *mtry*. The two parameters were permutated to form a set of parameter combinations. The performance of each parameter combination was evaluated using the RMSE. The parameter combination with the smallest RMSE was used to construct the final prediction model. For information on the optimized *ntree* and *mtry* for predicting RRFs, see Table S3. In the final model prediction step, the optimized MDs (predictors) and RF parameters were used to predict the RRTs or RRFs of the MeO-PCBs with the RFR models.

Model validation. The MLR and RFR models were validated with external datasets containing 12 MeO-PCBs for RRTs predictions and 11 MeO-PCBs for RRFs predictions (data for one MeO-PCBs was removed because it was below the detection limit) (see Table S1).

Candidate ranking in identifying unknown OH-PCBs (as methylated derivatives). Preliminary data analysis suggested that MeO-PCB isomers (i.e., varied chlorine or methoxy substitution patterns) have drastically different responses for the same MRM transition in the GC-MS/MS analysis (Fig. S2). Therefore, in addition to the predicted RRT, we used the predicted MS/MS data, consisting of the relative intensities of five fragment ions, to rank MeO-PCBs isomers derived from the same PCB congener or homolog to identify OH-PCBs in animals samples (i.e., feces and liver). For more details regarding the candidate ranking strategy, see the Supporting Information.

RESULTS AND DISCUSSION

Prediction of RRTs of MeO-PCBs. The identification of OH-PCBs in environmental and biological samples is challenging because of the large number of possible OH-PCBs and the structural similarity of OH-PCB metabolites of a specific PCB congener (e.g., PCB 95 or PCB 28). Therefore, it is unlikely that a single approach can achieve unambiguous identification of specific OH-PCB isomers; however, machine learning methods have the potential to aid in the identification of OH-PCB isomers.

We developed MLR and RFR models to predict the RRTs of MeO-PCBs on a GC-MS/MS system equipped with an SPB-Octyl column. Both models provide good approximations of the RRTs of MeO-PCBs, with R^2 values (derived from linear regressions between the measured and predicted values) greater than 0.98 (Fig. 1a) and with randomly distributed residuals (Fig. S3).

197 The MLR model with 10 predictors performed better, with a narrower prediction interval and
198 lower RMSE, than the RFR models with the same number of predictors. The absolute difference
199 between measured and predicted retention times was within 1 min for 87 % observations (n=121)
200 in the MLR model predictions. This finding is not surprising because statistically significant
201 linear relationships can be readily established between the predictors and the RRTs of MeO-
202 PCBs in the MLR development, with $p < 0.05$ for all 10 predictors (Table S2).
203 The MLR models developed with data from the SPB-octyl column slightly underestimate the
204 RRTs of MeO-PCBs collected with a different GC column (DB-1701) by overall 2 % (Fig. S4,
205 data was collected in a previous study), indicating a likely column flexibility, at least for poly(n-
206 octyl/methyl siloxane) phase columns. In addition to predicting the RRTs of MeO-PCBs, the
207 MLR models can also provide reasonable estimates of the RRTs of PCBs collected under
208 identical conditions but with a physically different instrument (Fig. S5). This finding indicates
209 that slight changes in chemical structure (e.g., with or without the methoxy group) and a
210 physically different instrument are unlikely to affect the model applications. However, the same
211 commercially available internal standards and similar instrument conditions are recommended to
212 apply the models to other problems. MLR models performed better than analogous RFR models
213 for the prediction of RRTs of MeO-PCBs on a DB-1701 column and RRTs of PCBs on an SPB-
214 Octyl column (Fig. S5).
215 This study is the first report of predictive models for OH-PCBs, but both MLR and RFR models
216 are widely used for predicting the retention times of chemicals on GC or LC systems. For
217 example, an MLR model with five PCB molecular descriptors (selected from topological
218 descriptors, geometric descriptors, electronic descriptors, and calculated physical property
219 descriptors) predicted the RRT of PCBs on a GC column with a relative standard deviation of 1.7

%.[45] Analogously, a five-variable MLR model with molecular electronegativity distance vectors of PCBs predicted the RRT of the PCBs with an RMSE of 0.0152 (or an MAE of approximately 1.90 min in retention time).[44] Retention times of chemicals were also predicted with RFR models on LC columns to facilitate the identification of unidentified peaks in untargeted metabolomics, with MAEs of 0.78 min (20 % in mean relative error) and 0.57 min (13 % in mean relative error) for hydrophilic interaction chromatography and reverse-phase LC columns, respectively.[47] The retention times of polybrominated diphenyl ethers and their methoxylated metabolites on a GC column were predicted with a lower accuracy by linear regression with the melting points.[46] Our MLR model with 10 predictors obtained comparable accuracy as above in predicting retention times of MeO-PCBs with an overall MAE of 0.55 min (n=121) (Fig. S3). However, the accuracy of the RRT predictions with this and other models does not meet the RRT variation tolerance recommended by the European Commission for identifying chromatographic peaks (i.e., 0.5 % and 2.5 % for GC and LC peaks, respectively).[63] Therefore, other identifiers, such as MS/MS profiles, are needed to identify unknown peaks.

Prediction of MS/MS profiles of MeO-PCBs. Principal component analysis and a violin plot of the MS/MS profiles of 99 mono- or di-MeO-PCBs suggested that their MS/MS data vary significantly with the position (i.e., ortho, meta, or para) of the methoxy group on the biphenyl moiety (Figs. 2a and S2). Notably, higher signals were observed for the loss of 50 (i.e., [CH_3+Cl]) for MeO-PCBs with ortho methoxy groups. On the other hand, meta- or para-methoxylated PCBs are more likely to fragment with the loss of 43 [CH_3+CO]. Since the loss of [CO] requires the opening of the MeO-substituted benzene ring, it is likely that the meta- and para-methoxylated PCBs chemically have a more favorable configuration for ring opening than

that of ortho-methoxylated PCBs, as illustrated in Fig. S6. This substitution pattern-dependent response suggests that MS/MS data can be used to assign the structure (i.e., ortho vs. meta or para-methoxy) of an unknown peak. Likewise, MS/MS responses were previously used to identify MeO-PCB 28 isomers formed in rats exposed to PCB 28.[64]

We predicted the MS/MS data of MeO-PCBs (expressed as the relative levels of the signals of the five fragmentations investigated) using RFR models coupled with MDs as predictors. The prediction of the RFR model, but not the MLR model, provided good approximations of the response for all five fragmentations, with MAE ranging from 0.3 to 0.5 log units (Figs. 2b-f). However, better estimations with a narrower prediction interval and lower MAE were obtained when predicting the RRFs associated with the loss of 43 or 50, likely because MeO-PCBs have higher responses generated through these two fragmentations. Importantly, the predicted MS/MS profiles were similar to the experimental data, with the similarity coefficient[65] $\cos\theta > 0.95$ for 92 % of the 96 MeO-PCBs investigated (Fig. 2g) ($\cos\theta = 1$ indicates that the MS/MS profiles are an exact match, $\cos\theta = 0$ indicates different profiles).

Since MS/MS data carry fragment information that can be used to identify unknown peaks, several programs (e.g., MetFrag,[66] CFM-ID,[48] and CSI:FingerID[67]) have been developed to predict the MS/MS data from the corresponding molecular structure. These programs were primarily designed for soft ionization systems, such as electrospray ionization (ESI), and provide no meaningful intensity values for the fragmentation of MeO-PCBs on a GC-MS/MS system with electron ionization (EI). Thus, the information provided by these software packages does not facilitate the identification of MeO-PCB isomers. CFM-ID has the option to simulate EI-MS spectra, but not EI-MS/MS spectra. Consequently, the intensity information predicted by this

approach in either EI-MS or ESI-MS/MS mode poorly reflects the experimental EI-MS/MS intensities in part because the CFM-ID program was originally not trained with reference MS spectra of MeO-PCBs (Fig. S7). Our machine-learning models were trained and externally validated with experimental MS/MS data of 124 mono/di-MeO-PCBs and, for the first time, allow the quantitative prediction of the MS/MS data of MeO-PCBs for which no authentic analytical standards are available. The predicted MS/MS data provide an additional dimension assisting in the identification of unknown MeO-PCB peaks.

Quantification of MeO-PCBs with the predicted RRFs. After the structural identification of an unknown MeO-PCB with the predicted retention time and MS/MS data, the unknown peak can be quantified with predicted RRFs. Since the MS/MS responses of MeO-PCBs depend on the position of the methoxy group on the biphenyl moiety (Fig. S2), we used signals of the respective transitions for the loss of 50 [CH_3+Cl] to quantify ortho-methoxylated PCBs and the loss of 43 [CH_3+CO] to quantify meta- or para-methoxylated PCBs. The levels of 89 MeO-PCBs (di-MeO-PCBs with both ortho- and meta/para-methoxy groups were excluded) predicted with this approach were within a 2-fold difference for 85 % observations and within a 3-fold difference for all observations (Fig. 3). These results demonstrate that the predicted RRFs allow a good approximation of the levels of OH-PCBs (as methylated derivatives) within one order of magnitude.

The RRFs of mono-MeO-PCBs for GC-MS analyses in the selected ion monitoring (SIM) mode have been predicted with a quadratic model using the number of chlorine atoms as a predictor.[35] This model was trained with one of the standard mixtures (Solution 1) used in this study (Fig. 3). The RRFs predicted by the quadratic model were verified by quantifying 12

mono-MeO-PCBs with values ranging from 0.8 to 2 times of the actual concentrations. The RRFs predicted by our RFR model estimated the levels of 96 % of the Solution 1 authentic analytical standards (n = 54, coeluting and di-MeO-PCBs were not included) within a 2-fold difference (0.5 – 2 times of the actual concentrations) and, thus, have similar accuracy as the earlier model. This observation is not surprising because the use of MRM signals increases the complexity of the modeling while increasing the selectivity in identifying unknowns. A lower accuracy was observed when estimating the levels of the second standard solution (Solution 2), likely because this standard solution contained most of the di-MeO-PCBs included in this study.

Characterization of OH-PCBs using predicted RRTs, MS/MS data, and RRFs. The flow chart in Fig. 4 illustrates how we propose to use the predicted RRT and MS/MS data to aid in the identification and quantification of OH-PCB metabolites (as methylated derivatives) in environmental or biological samples. Step 1: Sample extracts containing OH-PCBs are derivatized and analyzed by GC-MS/MS, as described in the Experimental Section, to collect experimental RRT and MS/MS data of the OH-PCBs. Step 2: For each OH-PCB metabolite peak, the RRTs of all possible MeO-PCB derivatives, as their SMILES structures, are predicted with our RRT prediction model. Step 3: The MS/MS data of all possible structures of an OH-PCB metabolite peak, also as their SMILES structures, are predicted with our MS/MS prediction model. Step 4: The weighted rank scores of all candidate structures are calculated (see the Supporting Information). Step 5: Identify the OH-PCB metabolite peaks based on the weighted rank scores. If available, a small set of MeO-PCB standards can be used to assist with the identification of the OH-PCB isomers. Step 6: The OH-PCB peaks are integrated and quantified using the predicted MS/MS responses. An additional dataset containing the detailed user manual

of these steps, example data and the R codes were publicly available in Iowa Research Online at http://doi.org/. The following section demonstrates the application of this approach to facilitate the identification and quantification of OH-PCB 95 in mouse feces and OH-PCB 28 in mouse liver. Since the model predictions were originally trained using experimental data obtained with standard solutions, these predictions facilitate the availability of standard retention times and MS/MS response factors independent of the sample matrix. OH-PCBs in any sample matrix can be theoretically identified and quantified with the predicted standard retention times and MS/MS data as long as necessary sample preparation procedures were performed, as described in this and other studies.[4, 14, 15, 58]

Analysis of OH-PCB 95 in the feces of a mouse exposed to PCB 95. PCB 95 and its metabolites are potentially neurotoxic.[49-52] Because metabolites of higher chlorinated PCBs are excreted with the feces,[68] we investigated OH-PCBs in a feces sample from a mouse exposed to PCB 95. We detected 5 peaks (Peaks 1, 2, 3, 4, and 5) with the MS transition *m/z* 356→313, corresponding to pentachlorinated mono-MeO-PCBs, and 2 peaks (Peaks 6 and 7) with the MS transition *m/z* 386→343, corresponding to pentachlorinated di-MeO-PCBs, in the extract of feces from a mouse exposed to PCB 95 (Fig. 5a). The possible mono-MeO-PCB 95 and selected di-MeO-PCB 95 that are likely formed in PCB metabolism studies, for example, metabolites with two methoxy groups ortho or para to each other, are shown in Fig. S8. The MeO-PCB 95 candidates were ranked based on their weighted rank scores calculated from the predicted and experimental RRT and MS/MS data (Fig. 5b).

Overall, the model correctly suggested the position of methoxy groups (ortho, meta, or para). Briefly, Peaks 1, 4, and 7 were correctly identified based on the weighted ranking scores as 3-103

(1,2-shift product), 4′-95, and 4,5-95, respectively. The weighted ranking scores suggested that Peaks 3 and 5 correspond to a meta- and para-hydroxylated metabolite (3′-95 and 4′-95, respectively). Based on the elution order of authentic analytical standards of MeO-PCB 95 analyzed on the same GC column (SPB-Octyl) (Fig. S9), Peaks 3 and 5 correspond to meta- and para-hydroxylate metabolites (5-95 and 4-95, respectively). These two correct identifications ranked within the top 3 candidates (Fig. 5b). Peak 2 was predicted to be 3′-95. This structural assignment requires confirmation with an analytical standard.

Peak 7 was correctly identified by the weighted rank scores as 4,5-PCB 95. The model also identified Peak 6 as 4,5-95, another catechol metabolite; however, Peak 6 likely corresponds to a different catechol metabolite, 3′,4′-95, as suggested by the top 2 candidate. This identification is consistent with the preferential formation of PCB catechol metabolites in PCB metabolism studies.[23-25] Finally, PCB 95 metabolites were quantified with their predicted RRFs. The predicted and experimental levels of the metabolites with available authentic standards (i.e., Peaks 1, 4, and 7) showed good agreement (Fig. 5c). Thus, the predicted RRF allows a reasonable approximation of the levels of PCB 95 metabolites for which no authentic analytical standards are available. The MS/MS responses of authentic standards of 5-95 (Peak 3) and 4-95 (Peak 5) were measured with a different GC-MS/MS method and were not included in the comparisons with the predicted levels in Fig. 5c.

The identification of PCB 95 metabolites using our model in combination with authentic analytical standards increases the confidence in the identification of unknown OH-PCB 95 metabolites in the feces sample from this study, but also earlier studies investigating the metabolism of PCB 95. For example, an unknown MeO-PCB 95 peak was detected in

metabolism studies with rat cytochrome P450 enzymes,[39] rat and human liver microsomes[36, 41] and in vivo disposition studies in rodent models.[37, 38, 40] In these previous studies we tentatively identified this unknown peak, which eluted before 5-95 on an SPB-1 column, as 3′-95. Our present study confirms this tentative identification of 3′-95 despite the difference in GC column stationary phases. Similarly, earlier metabolism studies with human liver microsomes or rats *in vivo* reported an unknown dihydroxylated PCB 95 metabolite peak (as its methylated derivative) that eluted before 4,5-95 on the SPB-1 column.[36, 37] In the absence of an authentic standard, the model predictions provide an additional line of evidence supporting the identification of this metabolite as 3′,4′-95, another PCB 95 catechol metabolite.

Analysis of OH-PCB 28 in the liver of a mouse exposed to a neurotoxic PCB mixture. We also investigated metabolites of PCB 28 in the liver from a mouse exposed during gestation and lactation to a PCB mixture.[53-55] Based on the MS transition *m/z* 286→243, we identified three trichlorinated MeO-PCB peaks (Peaks 1, 2, and 3) corresponding to mono-hydroxylated metabolites of PCB 28 (Fig. 6a). Based on the experimental and predicted RRT and MS/MS data, the weighted rank scores of all possible MeO-PCB 28 candidates (Fig. S10) were calculated for the three MeO-PCB 28 peaks (Fig. 6b). The top candidates for Peaks 1, 2, and 3 were 3′-28, 5-28, and 4-22 (a 1,2-shift product of PCB 28), respectively. The identification of Peaks 1 and 2 was subsequently confirmed with authentic standards. Using a small set of MeO-PCB 28 standards, we confirmed that Peak 3 does not correspond to 2′-28, 3-28, or 4′-25 (another 1,2-shift product of PCB28). Likely, Peak 3 was correctly identified as 4-22 by our model; however, confirmation with an authentic standard is still needed if this minor metabolite becomes a concern. The three peaks of PCB 28 metabolites were quantified with their predicted RRFs. As

374 with the PCB 95 metabolites above, the OH-PCB levels calculated with the predicted RRFs are
375 in good agreement with the experimental levels of the two metabolites for which authentic
376 analytical standards are available (i.e., 3′-28 and 5-28) (Fig. 6c).

377 Our predictions also enable a tentative identification of unknown metabolites observed in an
378 earlier study. Briefly, two major, meta-hydroxylated PCB 28 metabolites and two minor para-
379 hydroxylated PCB 28 metabolites (analyzed as methylated derivatives) were eliminated with the
380 feces of rats exposed intraperitoneally to PCB 28.[64] One meta-hydroxylated PCB 28 metabolite
381 was identified as 5-28 with a synthetic standard on a GC-MS equipped with a BP-5 column. The
382 other unidentified, meta-hydroxylated metabolite eluted at an earlier retention time. Based on the
383 elution order, we hypothesize that this metabolite corresponds to 3′-28 (Peak 1) observed in this
384 study (Fig. 6a), irrespective of the different GC columns used. The two para-hydroxylated PCB
385 28 metabolites were 1,2 shift products and remain unidentified because of the lack of analytical
386 standards. Similar to this study, one of the unknown para-hydroxylated PCB 28 metabolites
387 likely is 4-22.

388 The PCB metabolism studies described above highlight the complexity of the metabolism of
389 PCBs and the challenges associated with the identification of the PCB metabolites, which depend
390 on the availability of authentic analytical standards. The proposed strategy using machine
391 learning-based model predictions can significantly advance identifying and quantifying unknown
392 OH-PCBs, especially in combination with a small set of authentic analytical standards. Notably,
393 the predicted top candidate can suggest if the methoxy group is in the ortho, meta, or para
394 position. Even if the top candidate is not the true compound, knowing the position of the
395 methoxy substituent enables a targeted synthesis of authentic analytical standards. Additional

studies are needed to demonstrate that our machine learning approach can facilitate the identification of OH-PCB metabolites in environmental and biological samples.

CONFLICT OF INTEREST STATEMENT

The authors declare no competing financial interest.

FUNDING SOURCES

This work was supported by grants ES005605, ES013661, ES014901, ES027169, and ES031098 from the National Institute of Environmental Health Sciences, National Institutes of Health. The content is solely the responsibility of the authors. It does not necessarily represent the official views of the National Institute of Environmental Health Sciences, National Institutes of Health.

ACKNOWLEDGMENTS

Thanks to Drs. Ram Dhakal, Sudhir Joshi, and Sandhya Vyas from the University of Iowa for the synthesis of authentic analytical standards.

SUPPORTING INFORMATION

Sources and structures of MeO-PCBs, GC-MS/MS parameters, the candidate ranking algorithm, details regarding sample collection and extraction from animal experiments, optimal predictors and their linear coefficients and p-values to RRTs and MS/MS data of MeO-PCBs, diagnostic plots for model development, predominant pathways and data of the MeO-PCB fragmentations, model predictions of RRTs collected with a different GC column and RRTs of PCBs, CFM-ID predictions, and the structures, abbreviations and SMILES structures of MeO-

PCB 95 and MeO-PCB 28. This material is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES

1. Mandalakis, M.; Berresheim, H.; Stephanou, E. G. Direct evidence for destruction of polychlorobiphenyls by OH radicals in the subtropical troposphere. *Environ. Sci. Technol.* **2003,** *37*, 542-547.
2. Anderson, P. N.; Hites, R. A. OH radical reactions: The major removal pathway for polychlorinated biphenyls from the atmosphere. *Environ. Sci. Technol.* **1996,** *30*, 1756-1763.
3. Grimm, F. A.; Hu, D. F.; Kania-Korwel, I.; Lehmler, H. J.; Ludewig, G.; Hornbuckle, K. C.; Duffel, M. W.; Bergman, A.; Robertson, L. W. Metabolism and metabolites of polychlorinated biphenyls. *Crit. Rev. Toxicol.* **2015,** *45*, 245-272.
4. Marek, R. F.; Martinez, A.; Hornbuckle, K. C. Discovery of hydroxylated polychlorinated biphenyls (OH-PCBs) in sediment from a lake Michigan waterway and original commercial Aroclors. *Environ. Sci. Technol.* **2013,** *47*, 8204-8210.
5. Sari, M. F.; Esen, F.; Del Aguila, D. A. C.; Karakus, P. B. K. Passive sampler derived polychlorinated biphenyls (PCBs) in indoor and outdoor air in Bursa, Turkey: Levels and an assessment of human exposure via inhalation. *Atmos. Pollut. Res.* **2020,** *11*, 71-80.
6. Saktrakulkla, P.; Lan, T.; Hua, J.; Marek, R. F.; Thorne, P. S.; Hornbuckle, K. C. Polychlorinated biphenyls in food. *Environ. Sci. Technol.* **2020,** *54*, 11443-11452.
7. Jafarabadi, A. R.; Bakhtiari, A. R.; Mitra, S.; Maisano, M.; Cappello, T.; Jadot, C. First polychlorinated biphenyls (PCBs) monitoring in seawater, surface sediments and marine fish communities of the Persian Gulf: Distribution, levels, congener profile and health risk assessment. *Environ. Pollut.* **2019,** *253*, 78-88.

8. Sethi, S.; Keil, K. P.; Chen, H.; Hayakawa, K.; Li, X. S.; Lin, Y. P.; Lehmler, H. J.; Puschner, B.; Lein, P. J. Detection of 3,3 '-dichlorobiphenyl in human maternal plasma and its effects on axonal and dendritic growth in primary rat neurons. *Toxicol. Sci.* **2017,** *158*, 401-411.
9. Schettgen, T.; Esser, A.; Kraus, T.; Ziegler, P. Plasma levels of unintentionally produced non-Aroclor polychlorinated biphenyl (PCB) congeners in workers from the silicone rubber industry. *Chemosphere* **2022,** *291*, 132722.
10. Herkert, N. J.; Jahnke, J. C.; Hornbuckle, K. C. Emissions of tetrachlorobiphenyls (PCBs 47, 51, and 68) from polymer resin on kitchen cabinets as a non-aroclor source to residential air. *Environ. Sci. Technol.* **2018,** *52*, 5154-5160.
11. Anezaki, K.; Nakano, T. Concentration levels and congener profiles of polychlorinated biphenyls, pentachlorobenzene, and hexachlorobenzene in commercial pigments. *Environ. Sci. Pollut. Res.* **2014,** *21*, 998-1009.
12. Hu, D. F.; Hornbuckle, K. C. Inadvertent polychlorinated biphenyls in commercial paint pigments. *Environ. Sci. Technol.* **2010,** *44*, 2822-2827.
13. Hombrecher, K.; Quass, U.; Leisner, J.; Wichert, M. Significant release of unintentionally produced non-Aroclor polychlorinated biphenyl (PCB) congeners PCB 47, PCB 51 and PCB 68 from a silicone rubber production site in North Rhine-Westphalia, Germany. *Chemosphere* **2021,** *285*, 131449.
14. Marek, R. F.; Thome, P. S.; Herkert, N. J.; Awad, A. M.; Hornbuckle, K. C. Airborne PCBs and OH-PCBs inside and outside urban and rural US schools. *Environ. Sci. Technol.* **2017,** *51*, 7853-7860.

15. Marek, R. F.; Thorne, P. S.; Wang, K.; DeWall, J.; Hornbuckle, K. C. PCBs and OH-PCBs in serum from children and mothers in urban and rural U.S. communities. *Environ. Sci. Technol.* **2013,** *47*, 9555-9556.

16. Kawano, M.; Hasegawa, J.; Enomoto, T.; Ohishi, H.; Nishio, Y.; Matsuda, M.; Wakimoto, T. Hydroxylated polychlorinated biphenyls (OH-PCBs): recent advances in wildlife contamination study. *Environ. Sci.* **2005,** *12*, 315-324.

17. Pencikova, K.; Svrzkova, L.; Strapacova, S.; Neca, J.; Bartonkova, I.; Dvorak, Z.; Hyzdalova, M.; Pivnicka, J.; Palkova, L.; Lehmler, H. J.; Li, X. S.; Vondracek, J.; Machala, M. In vitro profiling of toxic effects of prominent environmental lower-chlorinated PCB congeners linked with endocrine disruption and tumor promotion. *Environ. Pollut.* **2018,** *237*, 473-486.

18. Machala, M.; Blaha, L.; Lehmler, H. J.; Pliskova, M.; Majkova, Z.; Kapplova, P.; Sovadinova, I.; Vondracek, J.; Malmberg, T.; Robertson, L. W. Toxicity of hydroxylated and quinoid PCB metabolites: Inhibition of gap junctional intercellular communication and activation of aryl hydrocarbon and estrogen receptors in hepatic and mammary cells. *Chem. Res. Toxicol.* **2004,** *17*, 340-347.

19. Grimm, F. A.; Lehmler, H. J.; He, X. R.; Robertson, L. W.; Duffel, M. W. Sulfated metabolites of polychlorinated biphenyls are high-affinity ligands for the thyroid hormone transport protein transthyretin. *Environ. Health Perspect.* **2013,** *121*, 657-662.

20. Ptak, A.; Ludewig, G.; Lehmler, H. J.; Wojtowicz, A. K.; Robertson, L. W.; Gregoraszczuk, E. L. Comparison of the actions of 4-chlorobiphenyl and its hydroxylated metabolites on

estradiol secretion by ovarian follicles in primary cells in culture. *Reprod. Toxicol.* **2005,** *20*, 57-64.

21. Pliskova, M.; Vondracek, J.; Canton, R. F.; Nera, J.; Kocan, A.; Petrik, J.; Trnovec, T.; Sanderson, T.; van den Berg, M.; Machala, M. Impact of polychlorinated biphenyls contamination on estrogenic activity in human male serum. *Environ. Health Perspect.* **2005,** *113*, 1277-1284.

22. Kester, M. H. A.; Bulduk, S.; Tibboel, D.; Meinl, W.; Glatt, H.; Falany, C. N.; Coughtrie, M. W. H.; Bergman, A.; Safe, S. H.; Kuiper, G. G. J. M.; Schuur, A. G.; Brouwer, A.; Visser, T. J. Potent inhibition of estrogen sulfotransferase by hydroxylated PCB metabolites: A novel pathway explaining the estrogenic activity of PCBs. *Endocrinology* **2000,** *141*, 1897-1900.

23. Zhang, C.-Y.; Flor, S.; Ruiz, P.; Dhakal, R.; Hu, X.; Teesch, L. M.; Ludewig, G.; Lehmler, H.-J. 3,3'-Dichlorobiphenyl is metabolized to a complex mixture of oxidative metabolites, including novel methoxylated metabolites, by HepG2 cells. *Environ. Sci. Technol.* **2020,** *54*, 12345-12357.

24. Dhakal, K.; Uwimana, E.; Adamcakova-Dodd, A.; Thorne, P. S.; Lehmler, H. J.; Robertson, L. W. Disposition of phenolic and sulfated metabolites after inhalation exposure to 4-chlorobiphenyl (PCB3) in female rats. *Chem. Res. Toxicol.* **2014,** *27*, 1411-1420.

25. McLean, M. R.; Bauer, U.; Amaro, A. R.; Robertson, L. W. Identification of catechol and hydroquinone metabolites of 4-monochlorobiphenyl. *Chem. Res. Toxicol.* **1996,** *9*, 158-164.

26. Spencer, W. A.; Lehmler, H. J.; Robertson, L. W.; Gupta, R. C. Oxidative DNA adducts after Cu2+-mediated activation of dihydroxy PCBs: Role of reactive oxygen species. *Free Radic. Biol. Med.* **2009,** *46*, 1346-1352.

27. Srinivasan, A.; Lehmler, H. J.; Robertson, L. W.; Ludewig, G. Production of DNA strand breaks in vitro and reactive oxygen species in vitro and in HL-60 cells by PCB metabolites. *Toxicol. Sci.* **2001,** *60*, 92-102.

28. Lin, P. H.; Sangaiah, R.; Ranasinghe, A.; Upton, P. B.; La, D. K.; Gold, A.; Swenberg, J. A. Formation of quinonoid-derived protein adducts in the liver and brain of Sprague-Dawley rats treated with 2,2 ',5,5 '-tetrachlorobiphenyl. *Chem. Res. Toxicol.* **2000,** *13*, 710-718.

29. Amaro, A. R.; Oakley, G. G.; Bauer, U.; Spielmann, H. P.; Robertson, L. W. Metabolic activation of PCBs to quinones: Reactivity toward nitrogen and sulfur nucleophiles and influence of superoxide dismutase. *Chem. Res. Toxicol.* **1996,** *9*, 623-629.

30. Espandiari, P.; Glauert, H. P.; Lehmler, H. J.; Lee, E. Y.; Srinivasan, C.; Robertson, L. W. Initiating activity of 4-chlorobiphenyl metabolites in the resistant hepatocyte model. *Toxicol. Sci.* **2004,** *79*, 41-46.

31. Espandiari, P.; Glauert, H. P.; Lehmler, H. J.; Lee, E. Y.; Srinivasan, C.; Robertson, L. W. Polychlorinated biphenyls as initiators in liver carcinogenesis: resistant hepatocyte model. *Toxicol. Appl. Pharm.* **2003,** *186*, 55-62.

32. Zhang, C.-Y.; Flor, S.; Ludewig, G.; Lehmler, H.-J. Atropselective partitioning of polychlorinated biphenyls in a HepG2 cell culture system: experimental and modeling results. *Environ. Sci. Technol.* **2020,** *54*, 13817-13827.

33. Uwimana, E.; Ruiz, P.; Li, X. S.; Lehmler, H. J. Human CYP2A6, CYP2B6, AND CYP2E1 atropselectively metabolize polychlorinated biphenyls to hydroxylated metabolites. *Environ. Sci. Technol.* **2019,** *53*, 2114-2123.

34. Zhang, D.; Saktrakulkla, P.; Tuttle, K.; Marek, R. F.; Lehmler, H. J.; Wang, K.; Hornbuckle, K. C.; Duffel, M. W. Detection and quantification of polychlorinated biphenyl sulfates in human serum. *Environmental Science & Technology* **2021,** *55*, 2473-2481.

35. Saktrakulkla, P.; Dhakal, R. C.; Lehmler, H. J.; Hornbuckle, K. C. A semi-target analytical method for quantification of OH-PCBs in environmental samples. *Environ. Sci. Pollut. Res.* **2020,** *27*, 8859-8871.

36. Uwimana, E.; Li, X. S.; Lehmler, H. J. 2,2 ',3,5 ',6-Pentachlorobiphenyl (PCB 95) is atropselectively metabolized to para-hydroxylated metabolites by human liver microsomes. *Chem. Res. Toxicol.* **2016,** *29*, 2108-2110.

37. Stamou, M.; Uwimana, E.; Flannery, B. M.; Kania-Korwel, I.; Lehmler, H. J.; Lein, P. J. Subacute nicotine co-exposure has no effect on 2,2 ',3,5 ',6-pentachlorobiphenyl disposition but alters hepatic cytochrome P450 expression in the male rat. *Toxicology* **2015,** *338*, 59-68.

38. Kania-Korwel, I.; Barnhart, C. D.; Lein, P. J.; Lehmler, H. J. Effect of pregnancy on the disposition of 2,2 ',3,5 ',6-pentachlorobiphenyl (PCB 95) atropisomers and their hydroxylated metabolites in female mice. *Chem. Res. Toxicol.* **2015,** *28*, 1774-1783.

39. Lu, Z.; Kania-Korwel, I.; Lehmler, H. J.; Wong, C. S. Stereoselective formation of mono- and dihydroxylated polychlorinated biphenyls by rat cytochrome P450 2B1. *Environ. Sci. Technol.* **2013,** *47*, 12184-12192.

40. Kania-Korwel, I.; Barnhart, C. D.; Stamou, M.; Truong, K. M.; El-Komy, M. H.; Lein, P. J.; Veng-Pedersen, P.; Lehmler, H.-J. 2,2',3,5',6-Pentachlorobiphenyl (PCB 95) and its hydroxylated metabolites are enantiomerically enriched in female mice. *Environ. Sci. Technol.* **2012,** *46*, 11393-11401.

41. Kania-Korwel, I.; Duffel, M. W.; Lehmler, H. J. Gas chromatographic analysis with chiral cyclodextrin phases reveals the enantioselective formation of hydroxylated polychlorinated biphenyls by rat liver microsomes. *Environ. Sci. Technol.* **2011,** *45*, 9590-9596.

42. Saktrakulkla, P.; Li, X.; Martinez, A.; Lehmler, H.-J.; Hornbuckle, K. C. Hydroxylated Polychlorinated Biphenyls Are Emerging Legacy Pollutants in Contaminated Sediments. *Environ. Sci. Technol.* **2022,** *56*, 2269-2278.

43. Ueno, D.; Darling, C.; Alaee, M.; Campbell, L.; Pacepavicius, G.; Teixeira, C.; Muir, D. Detection of Hydroxylated Polychlorinated Biphenyls (OH-PCBs) in the Abiotic Environment: Surface Water and Precipitation from Ontario, Canada. *Environ. Sci. Technol.* **2007,** *41*, 1841-1848.

44. Liu, S. S.; Liu, Y.; Yin, D. Q.; Wang, X. D.; Wang, L. S. Prediction of chromatographic relative retention time of polychlorinated biphenyls from the molecular electronegativity distance vector. *J. Sep. Sci.* **2006,** *29*, 296-301.

45. Hasan, M. N.; Jurs, P. C. Computer-assisted prediction of gas-chromatographic retention times of polychlorinated-biphenyls. *Anal. Chem.* **1988,** *60*, 978-982.

46. Simpson, S.; Gross, M. S.; Olson, J. R.; Zurek, E.; Aga, D. S. Identification of polybrominated diphenyl ether metabolites based on calculated boiling points from

566 COSMO-RS, experimental retention times, and mass spectral fragmentation patterns. *Anal.*
567 *Chem.* **2015,** *87*, 2299-2305.
568 47. Bonini, P.; Kind, T.; Tsugawa, H.; Barupal, D. K.; Fiehn, O. Retip: Retention time
569 prediction for compound annotation in untargeted metabolomics. *Anal. Chem.* **2020,** *92*,
570 7515-7522.
571 48. Allen, F.; Pon, A.; Wilson, M.; Greiner, R.; Wishart, D. CFM-ID: a web server for
572 annotation, spectrum prediction and metabolite identification from tandem mass spectra.
573 *Nucleic Acids Res.* **2014,** *42*, W94-W99.
574 49. Niknam, Y.; Feng, W.; Cherednichenko, G.; Dong, Y.; Joshi, S. N.; Vyas, S. M.; Lehmler,
575 H.-J.; Pessah, I. N. Structure-Activity Relationship of Selected Meta- and Para-
576 Hydroxylated Non–Dioxin Like Polychlorinated Biphenyls: From Single RyR1 Channels to
577 Muscle Dysfunction. *Toxicol. Sci.* **2013,** *136*, 500-513.
578 50. Wayman Gary, A.; Yang, D.; Bose Diptiman, D.; Lesiak, A.; Ledoux, V.; Bruun, D.; Pessah
579 Isaac, N.; Lein Pamela, J. PCB-95 Promotes Dendritic Growth via Ryanodine Receptor–
580 Dependent Mechanisms. *Environ. Health Perspect.* **2012,** *120*, 997-1002.
581 51. Wayman Gary, A.; Bose Diptiman, D.; Yang, D.; Lesiak, A.; Bruun, D.; Impey, S.; Ledoux,
582 V.; Pessah Isaac, N.; Lein Pamela, J. PCB-95 Modulates the Calcium-Dependent Signaling
583 Pathway Responsible for Activity-Dependent Dendritic Growth. *Environ. Health Perspect.*
584 **2012,** *120*, 1003-1009.
585 52. Pessah, I. N.; Hansen, L. G.; Albertson, T. E.; Garner, C. E.; Ta, T. A.; Do, Z.; Kim, K. H.;
586 Wong, P. W. Structure−Activity Relationship for Noncoplanar Polychlorinated Biphenyl

587 Congeners toward the Ryanodine Receptor-Ca2+ Channel Complex Type 1 (RyR1). *Chem.*
588 *Res. Toxicol.* **2006,** *19*, 92-101.
589 53. Sethi, S.; Keil Stietz, K. P.; Valenzuela, A. E.; Klocke, C. R.; Silverman, J. L.; Puschner,
590 B.; Pessah, I. N.; Lein, P. J. Developmental Exposure to a Human-Relevant Polychlorinated
591 Biphenyl Mixture Causes Behavioral Phenotypes That Vary by Sex and Genotype in
592 Juvenile Mice Expressing Human Mutations That Modulate Neuronal Calcium. *Frontiers in*
593 *Neuroscience* **2021,** *15*, 766826.
594 54. Matelski, L.; Keil Stietz, K. P.; Sethi, S.; Taylor, S. L.; Van de Water, J.; Lein, P. J. The
595 influence of sex, genotype, and dose on serum and hippocampal cytokine levels in juvenile
596 mice developmentally exposed to a human-relevant mixture of polychlorinated biphenyls.
597 *Curr. Res. Toxicol.* **2020,** *1*, 85-103.
598 55. Rude, K. M.; Pusceddu, M. M.; Keogh, C. E.; Sladek, J. A.; Rabasa, G.; Miller, E. N.;
599 Sethi, S.; Keil, K. P.; Pessah, I. N.; Lein, P. J.; Gareau, M. G. Developmental exposure to
600 polychlorinated biphenyls (PCBs) in the maternal diet causes host-microbe defects in
601 weanling offspring mice. *Environ. Pollut.* **2019,** *253*, 708-721.
602 56. Egusquiza, R. J.; Ambrosio, M. E.; Wang, S. G.; Kay, K. M.; Zhang, C.; Lehmler, H.-J.;
603 Blumberg, B. Evaluating the Role of the Steroid and Xenobiotic Receptor (SXR/PXR) in
604 PCB-153 Metabolism and Protection against Associated Adverse Effects during Perinatal
605 and Chronic Exposure in Mice. *Environ. Health Perspect.* **2020,** *128*, 047011.
606 57. Wu, X. A.; Pramanik, A.; Duffel, M. W.; Hrycay, E. G.; Bandiera, S. M.; Lehmler, H. J.;
607 Kania-Korwel, I. 2,2 ',3,3 ',6,6 '-Hexachlorobiphenyl (PCB 136) is enantioselectively

oxidized to hydroxylated metabolites by rat Liver microsomes. *Chem. Res. Toxicol.* **2011,** *24*, 2249-2257.

58. Awad, A. M.; Martinez, A.; Marek, R. F.; Hornbuckle, K. C. Occurrence and distribution of two hydroxylated polychlorinated biphenyl congeners in Chicago air. *Environ. Sci. Technol. Lett.* **2016,** *3*, 47-51.

59. Rajarshi Guha; Charlop-Powers, Z.; Schymanski, E. rcdk: Interface to the 'CDK' Libraries. https://cran.r-project.org/web/packages/rcdk/index.html (Accessed on July 30, 2022),

60. Manly, B. F. J.; Alberto, J. A. N., *Multivariate Statistical Methods: A Primer*. 4th ed.; Chapman and Hall/CRC: New York, 2016; p 14.

61. Konovalov, D. A.; Llewellyn, L. E.; Heyden, Y. V.; Coomans, D. Robust cross-validation of linear regression QSAR models. *J. Chem. Inf. Model.* **2008,** *48*, 2081-2094.

62. Shao, J. Linear-model selection by cross-validation. *J Am Stat Assoc* **1993,** *88*, 486-494.

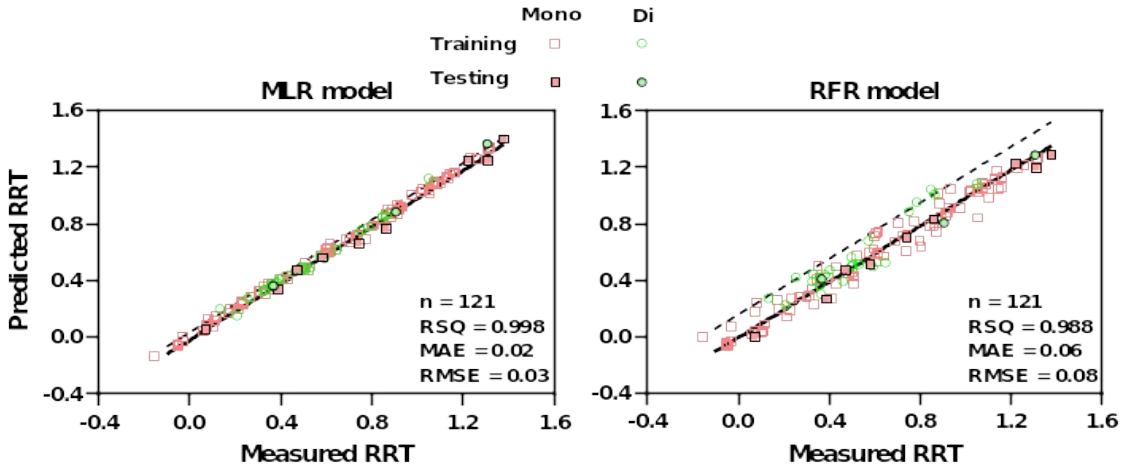
63. Commission, E. Commission Decision EC 2002/657 of 12 August 2002 implementing Council Directive 96/23/EC concerning the performance of analytical methods and the interpretation of results. *Off. J. Eur. Communities: Legis* **2002,** *L221*, 8-36.

64. Moir, D.; Viau, A.; Chu, I.; Wehler, E. K.; Morck, A.; Bergman, A. Tissue distribution, metabolism, and excretion of 2,4,4'-trichlorobiphenyl (CB-28) in the rat. *Toxicol. Ind. Health.* **1996,** *12*, 105-121.

65. Davis, J. C., *Statistics and data analysis in geology*. 3rd ed.; John Wiley & Sons: New York, 2002; p 540.

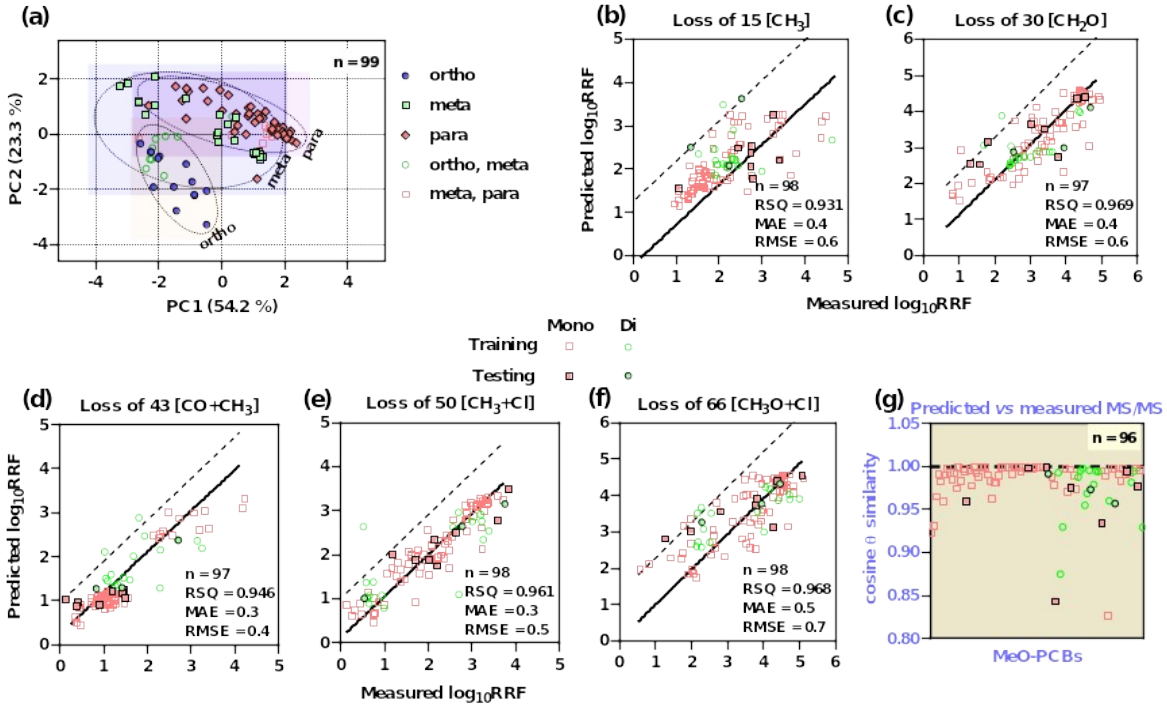
66. Wolf, S.; Schmidt, S.; Müller-Hannemann, M.; Neumann, S. In silico fragmentation for computer assisted identification of metabolite mass spectra. *BMC Bioinform.* **2010,** *11*, 148.

630 67. Dührkop, K.; Shen, H.; Meusel, M.; Rousu, J.; Böcker, S. Searching molecular structure
631 databases with tandem mass spectra using CSI:FingerID. *Proc. Natl. Acad. Sci. U.S.A.*
632 **2015,** *112*, 12580-12585.
633 68. Birnbaum, L. S. The Role of Structure in the Disposition of Halogenated Aromatic
634 Xenobiotics. *Environ. Health Perspect.* **1985,** *61*, 11-20.
635



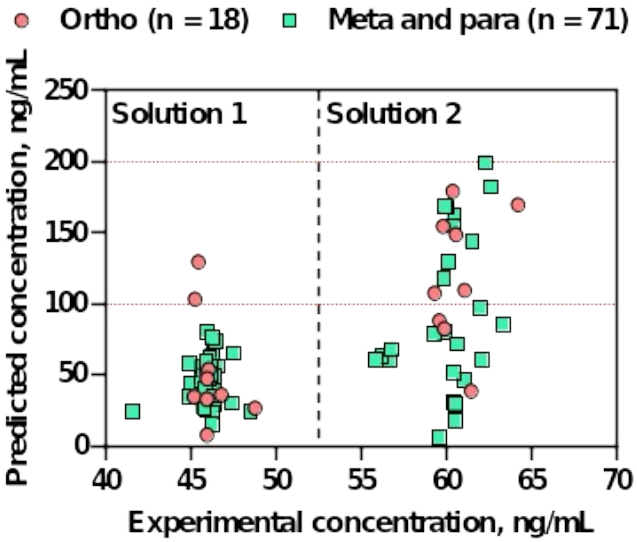
636

637 **Fig.1.** (a) A multiple linear regression (MLR) model provided a better estimation of the RRTs of
638 MeO-PCBs compared to the random forest regression (RFR) model. The model training datasets
639 were constructed with the measured RRTs and molecular descriptors of 87 mono-MeO-PCBs
640 and 22 di-MeO-PCBs. The testing dataset contains the measured RRTs and molecular
641 descriptors of nine mono-MeO-PCBs (mono- to nona-chlorinated) and three di-MeO-PCBs (di-,
642 tetra-, or octa-chlorinated). The dash lines in panel (a) indicate the borders of the prediction
643 interval with a 95 % confidence level.



644

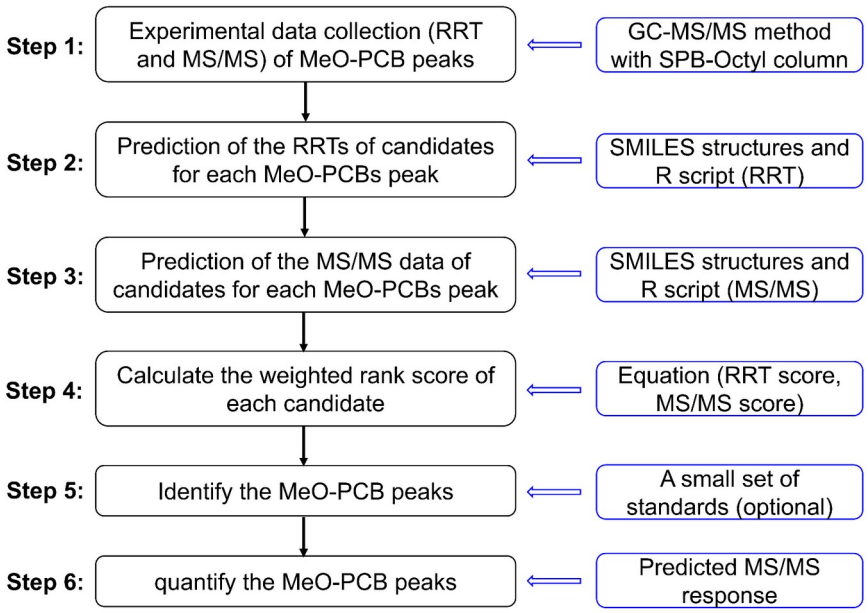
645 **Fig. 2.** The responses of five fragmentations (i.e., the loss of 15 [CH_3], 30 [CH_2O], 43
646 [CH_3+CO], 50 [CH_3+Cl] and 66 [CH_3O+Cl]) of the MeO-PCBs varied with the position (*ortho*,
647 *meta*, or *para*) of the methoxy group, as revealed by (a) a principal component analysis (PCA).
648 (b-f) Random forest regression model with molecular descriptors as predictors provided
649 reasonable estimations of the responses of five fragmentations studied. The model training and
650 testing datasets were constructed with the MS/MS data (expressed as the relative response
651 factors) from 88 and 11 observations, respectively. The dash lines indicate the borders of the
652 prediction interval with a 95 % confidence level. (g) The similarity coefficient cos θ showed
653 agreement between predicted and measured MS/MS profiles of MeO-PCBs.



654

655 **Fig. 3.** A comparison of the levels of MeO-PCBs quantified by predicted relative response
656 factors (RRFs) with experimental values. The RRFs of MeO-PCBs were predicted with the
657 random forest regression model coupled with the molecular structures. The ortho-methoxylated
658 PCBs were quantified with RRFs predicted for the loss of 50 [CH_3+Cl], and the meta- and para-
659 methoxylated PCBs were quantified with RRFs predicted for the loss of 43 [CH_3+CO]. Two
660 MeO-PCBs standard mixtures (Solution 1 and Solution 2) with concentrations of 47 and 60
661 ng/mL, respectively, were used.

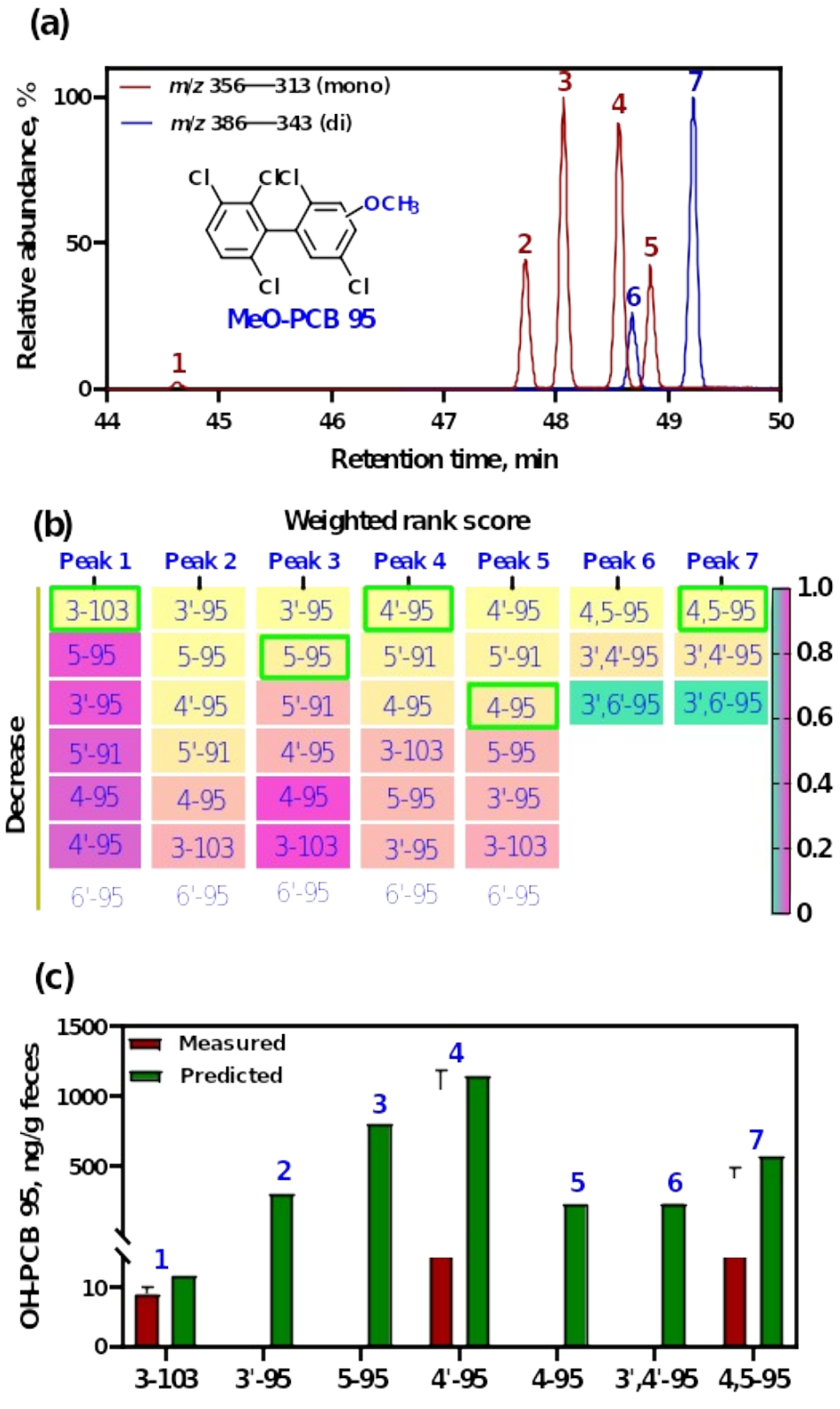
662



663

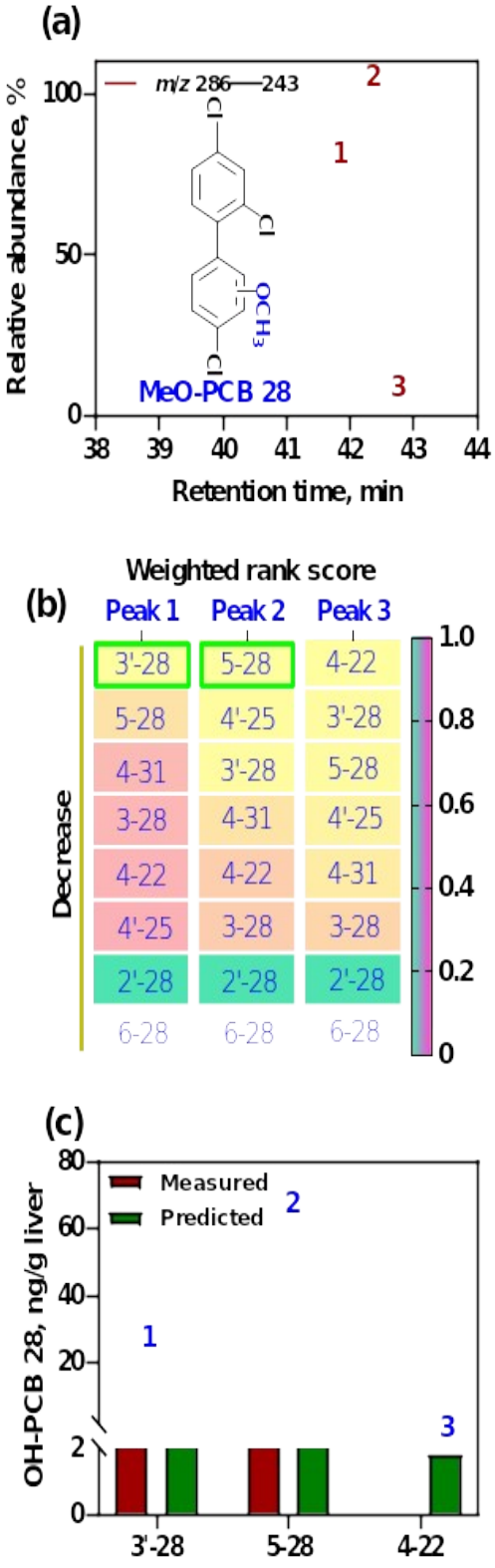
664 **Fig. 4**. Proposed workflow for the characterization and quantification of OH-PCBs (analyzed as

665 methylated derivatives) using predicted retention times (RRT) and MS/MS responses.


(a)
m/z 356—313 (mono)
m/z 386—343 (di)
Relative abundance, %
Retention time, min
MeO-PCB 95
(b)
Weighted rank score
Peak 1
Peak 2
Peak 3
Peak 4
Peak 5
Peak 6
Peak 7
Decrease
(c)
Measured
Predicted
OH-PCB 95, ng/g feces
3-103
3'-95
5-95
4'-95
4-95
3',4'-95
4,5-95

666

Fig. 5. (a) GC-MS/MS chromatograms indicate the presence of five peaks (Peaks 1, 2, 3, 4, and 5) of mono-hydroxylated metabolites and two peaks of di-hydroxylated metabolites (Peaks 6 and 7) in a feces sample from a mouse orally exposed to PCB 95. The OH-PCBs were analyzed as methylated derivatives. (b) Possible candidates for each peak were proposed and ranked based on their weighted scores calculated with measured and predicted retention times and MS/MS data. The candidate structures of OH-PCB in this and the following figures are abbreviated with the position of the OH group plus their PCB number, for example 4-95. The candidates in green borders was unambiguously identified with an authentic standard. (c) The agreement between measured and predicted levels of the OH-PCB 95 metabolites (i.e., 3-103, 4′-95 and 4,5-PCB 95) supports the quantification of OH-PCBs with a predicted relative response factor. The abbreviations and the corresponding structures of the MeO-PCB 95 metabolites are provided in Fig. S6.


(a)
m/z 286—243
2
1
3
MeO-PCB 28
Relative abundance, %
Retention time, min
(b)
Weighted rank score
Peak 1
Peak 2
Peak 3
Decrease
3'-28
5-28
4-22
5-28
4'-25
3'-28
4-31
3'-28
5-28
3-28
4-31
4'-25
4-22
4-22
4-31
4'-25
3-28
3-28
2'-28
2'-28
2'-28
2'-28
6-28
6-28
6-28
1.0
0.8
0.6
0.4
0.2
0
(c)
Measured
Predicted
2
1
3
OH-PCB 28, ng/g liver
3'-28
5-28
4-22

679

Fig. 6. (a) GC-MS/MS chromatograms support the formation of three peaks (Peak 1, 2, and 3) of mono-hydroxylated metabolites of PCB 28 in a liver sample collected from a mouse exposed throughout gestation and lactation to a PCB mixture (6 mg/kg/day) containing PCB 28 as a major component. (b) Possible candidate for each metabolite peak were propose and ranked with their weighted scores calculated with measured and predicted retention times and MS/MS data. The candidates in green borders were unambiguously identified with an authentic standard. (c) The agreement between measured and predicted levels of the OH-PCB 28 metabolites supports the quantification of OH-PCBs with a predicted relative response factor. The abbreviations and the corresponding structures of the MeO-PCB 28 metabolites are provided in Fig. S8.