Title

Modeling the transplacental transfer of small molecules using machine learning: a case study on per- and polyfluorinated substances (PFAS)

Journal

Journal of Exposure Science & Environmental Epidemiology, 32(6)

ISSN

1559-0631

Authors

Abrahamsson, Dimitri
Siddharth, Adi
Robinson, Joshua F
et al.

Publication Date

2022-11-01

DOI

10.1038/s41370-022-00481-2

Peer reviewed

HHS Public Access

Author manuscript

J Expo Sci Environ Epidemiol. Author manuscript; available in PMC 2023 April 07.

Published in final edited form as:

J Expo Sci Environ Epidemiol. 2022 November ; 32(6): 808–819. doi:10.1038/s41370-022-00481-2.

Modeling the transplacental transfer of small molecules using machine learning: A case study on per- and polyfluorinated substances (PFAS)

Dimitri Abrahamsson[*,1], **Adi Siddharth**[1], **Joshua F. Robinson**[1], **Anatoly Soshilov**[2], **Sarah Elmore**[2], **Vincent Cogliano**[2], **Carla Ng**[3], **Elaine Khan**[2], **Randolph Ashton**[4,5,6], **Weihsueh A. Chiu**[7], **Jennifer Fung**[8], **Lauren Zeise**[2], **Tracey J. Woodruff**[*,1]

[1]Department of Obstetrics, Gynecology and Reproductive Sciences, Program on Reproductive Health and the Environment, University of California, San Francisco, 490 Illinois Street, San Francisco, CA 94143, United States.

[2]California Environmental Protection Agency, Office for Environmental Health and Hazard Assessment, 1515 Clay St., Oakland, CA 94612, United States.

[3]Department of Civil and Environmental Engineering, University of Pittsburgh, 3700 O'Hara St, PA 15261, United States.

[4]Wisconsin Institute for Discovery, University of Wisconsin, Madison, 330 N Orchard St, Madison, WI 53715 Wisconsin

[5]The Stem Cell and Regenerative Medicine Center, University of Wisconsin, Madison, 1111 Highland Avenue, Madison, WI 53705 Wisconsin

[6]Department of Biomedical Engineering, University of Wisconsin - Madison, 1550 Engineering Drive, Madison, Wisconsin 53706, United States

[7]Department of Veterinary Integrative Biosciences, College of Veterinary Medicine and Biomedical Sciences, Texas A&M University, College Station, Texas 77843, United States.

[8]Department of Obstetrics, Gynecology, and Reproductive Science and the Center of Reproductive Science, University of California, San Francisco, San Francisco, CA 94143-2240, USA

Abstract

[*]**Corresponding authors**: Dimitri Abrahamsson, dimitri.abrahamsson@gmail.com and Tracey J. Woodruff, tracey.woodruff@ucsf.edu.

Author contributions

D.A. collected the data, conducted the analyses and wrote the manuscript. A.Si. assisted in the data collection and analysis. J.F.S., J.F., and T.J.W supervised, provided feedback and assisted with the writing of the manuscript. A.So., S.E., V.C., E.K., and L.Z. provided feedback on the analyses and assisted with the writing of the manuscript. C.N., R.A., W.C. constituted the advisory panel of the study, provided feedback and assisted with the analyses and the writing of the manuscript.

Conflict of interest

The authors declare that to the best of their knowledge they have no conflict of interest.

Background: Despite their large numbers and widespread use, very little is known about the extent to which per- and polyfluoroalkyl substances (PFAS) can cross the placenta and expose the developing fetus.

Objective: The aim of our study is to develop a computational approach that can be used to evaluate the of extend of which small molecules, and in particular PFAS, can cross to cross the placenta and partition to cord blood.

Methods: We collected experimental values of the central tendency of concentration ratio between cord and maternal blood (R_{CM}) for 260 chemical compounds and calculated their physicochemical descriptors using the cheminformatics package Mordred. We developed and tested an artificial neural network (ANN) and used the compiled database to train the model. We then applied our best performing model to make predictions of R_{CM} for a large dataset of PFAS chemicals (n=7,982). We, finally, used the calculated descriptors of the chemicals to identify which properties correlated significantly with R_{CM}.

Results: We determined that 7855 compounds were within the applicability domain and 127 compounds are outside the applicability domain of our model. Our predictions of R_{CM} for PFAS suggested that 3623 compounds had a log $R_{CM} > 0$ indicating preferable partitioning to cord blood. Some examples of these compounds were bisphenol AF, 2,2-bis(4-aminophenyl)hexafluoropropane and nonafluoro-tert-butyl 3-methylbutyrate.

Significance: These observations have important public health implications as many PFAS have been shown to interfere with fetal development.

1. Introduction

1.1 Chemicals and pregnancy

During pregnancy, the placenta regulates the transfer and exchange of endogenous molecules (e.g., hormones), nutrients, gases and waste from the mother to the fetus and vice versa.[1] As humans are exposed to environmental contaminants in their indoor and outdoor environment, these chemicals enter the human body and are transported to the various tissues as blood circulates through the body. While the placenta is often efficient at removing waste and toxicants from the fetus, many environmental contaminants can cross the placenta and expose the fetus.[2–4]

1.2 Partitioning between maternal and cord blood

Several studies have reported on the occurrence of environmental contaminants in maternal and cord blood[2,5–7], however, the mechanism by which chemicals cross or are prevented from crossing the placenta is not well understood. Transfer through passive diffusion is thought to be the driving force behind transplacental transfer, however, active transport facilitated diffusion, phagocytosis, and pinocytosis are also expected to play a role.[8] With an ever-increasing number of chemicals released to the U.S. market, there is a need to develop predictive models that can be used to calculate the extent to which chemicals can cross the placenta and expose the fetus. Such approaches will allow regulators and the scientific community to target compounds with a potential risk for adverse health effects in fetuses which can in turn lead to adverse health outcomes later in childhood and adulthood.

1.3 PFAS

One class of chemicals that has recently come under scrutiny are per- and polyfluoroalkyl substances (PFAS).[9–11] PFAS compose a diverse group of organofluorine chemicals with varying structures and functions. Although the exact number of PFAS currently in the environment is unknown, recent efforts to map the known chemicals in this class have led to a compiled list by the U.S. Environmental Protection Agency (EPA) that contains 9,252 chemicals.[12] The list ("PFASMASTER") is available on EPA's CompTox Chemicals Dashboard[12,13] (referred to as "the dashboard" from now on) and it is compilation of smaller lists from other agencies, such as the Swedish Chemicals Inspection Agency (KEMI).[14] It should be noted that while the number of PFAS was 9,252 at the time that this manuscript was written, this number is expected to change in the future as the dashboard is a dynamic project and the chemical lists are continuously updated. Many studies have demonstrated the potential for persistence[15], bioaccumulation[16] and toxicity[17] of several PFAS in the environment[18] and in humans[19], however, the vast majority of PFAS on PFASMASTER remain uncharacterized.

1.4 PFAS and human health

Exposures to PFAS has been shown to increase the risk of numerous health effects, such as immune suppression including decreased vaccine effectiveness[20,21], thyroid function[22], liver disease[23] and cancer[24], decrease in fertility[25] and lower birth weight.[26] PFAS exposures are widespread in the population, with biomonitoring studies reporting over 90% of the population exposed to one or more PFAS. Additionally, multiple studies report PFAS measurements in paired maternal and cord blood samples, demonstrating fetal exposures[6,27,28]. Many PFAS have been shown to cross the placenta and partition preferably to cord blood increasing the risk for adverse effects in the fetus. Some examples of PFAS that have shown to preferentially partition to cord blood are perfluorohexanoic acid (PFHxA)[4], perfluorobutanoic acid (PFBA)[29], and perfluorotetradecanoic acid (PFTeDA)[4]. Fetal exposures during pregnancy have been associated with low birth weight[30], dysregulated lipid metabolism in utero[31], adiposity in early childhood in girls,[32] and neuro-behavioral outcomes in early childhood, such as hyperactivity[33]. Understanding and predicting unknown fetal exposures is of crucial importance in order to protect fetal health and development.

Common approaches of predicting fetal exposures involve 1) physiologically based pharmacokinetic (PBPK) models[34,35] that describe the partitioning of chemicals in the various tissues of the mother and fetus and simpler 2) pharmacokinetic models[36,37] (PK) that may describe the partitioning between mother and fetus as two distinct compartments. While these approaches can be useful in developing a mechanistic understanding of the behavior of chemicals, they can often be restricted by the limited amount of data that can be used for validation, given the difficulties associated with biomonitoring of chemical compounds in fetal tissues. For example, for PFAS, there is only a small number of studies that have evaluated concentrations in fetal tissues[38,39] or in adult human tissues.[40] The limited number of studies is most likely due to the difficulty of obtaining and analyzing human fetal tissues. Other approaches involve describing the partitioning of chemicals between maternal and cord blood as a ratio of concentrations (R_{CM}) and/or using concentrations in cord blood

as a surrogate for fetal exposures. Contrary to reports of PFAS in fetal tissues, there is a substantially larger number of studies that report concentrations of PFAS in matched maternal and cord blood samples[3,4,27] and have evaluated factors that influence the transfer of PFAS from maternal to cord blood. Typically, such approaches use multilinear regressions built with physicochemical descriptors (e.g., length or carbon chain) and physiological descriptors (e.g., gestational diabetes) as a tool for describing and predicting the partitioning of chemical compounds between maternal and cord blood.[3,41] While these approaches can be informative when it comes to describing the partitioning of chemicals between maternal and cord blood, their predictive power is often limited.

Advances in machine learning have enabled us to improve the predictions of such properties, like R_{CM}, by utilizing advanced machine learning algorithms, such as random forest (RF), artificial neural network (ANN) and support vector machine (SVM), among others. Machine learning facilitates the integration of large number of parameters that can be important in the predictions and allow the model to determine the importance of each parameter. Improved predictions of R_{CM} are useful in characterizing fetal exposures and understanding the related health risks.

1.5 Our aim

The overall goal of our study was to develop and evaluate machine learning approaches to predict R_{CM} and identify physicochemical properties of PFAS that correlate with increased fetal exposures and can be used to predict fetal exposures based on maternal data. To accomplish this, we had four aims: 1) to compile a database with R_{CM} values of small molecules from the literature to use as a training set for machine learning; 2) to develop a predictive machine learning model that can be used to calculate the R_{CM}, 3) to apply the algorithm to predict R_{CM} to all PFAS whose physicochemical space overlaps with that of the chemicals in our compiled database; and 4) to understand the physicochemical properties and the structural characteristics that determine the partitioning of small molecules between cord and maternal blood.

2. Methods

2.1 Partitioning

The partitioning of a chemical between cord and maternal blood can be described as the ratio between cord and maternal blood concentrations (R_{CM}):

$$R_{CM} = \frac{[x]_C}{[x]_M}$$

Where, $[x]_C$ is the concentration of a chemical x in cord blood and $[x]_M$ is the concentration in maternal blood. It is important to note that R_{CM} does not describe an equilibrium partition ratio, such as the octanol-water equilibrium partition ratio (K_{OW}), but rather a concentration ratio representing the current state of a dynamic system. As the placenta is a dynamic system, through which, chemicals are actively flowing, it is unlikely that any chemicals will

be at equilibrium. The partitioning of chemicals between cord and maternal blood has also been described as a concentration ratio in previous studies.[28,42,43]

2.2 Collecting the data

Our workflow for collecting and compiling the database is presented as a flowchart in Fig. 1. As a first step, we searched the published literature for studies that measured paired maternal and cord blood or serum chemical concentrations in humans for small molecules (<1000 Da). The keywords we used to search for relevant studies were: [‘maternal’] AND [‘cord’] AND [‘chemical concentration’] OR [‘drug concentration’] OR [‘pharmaceutical concentration’] in the following databases: Web of Science, SciFinder, ACS Publications and Google Scholar. We selected studies that reported measurements for paired maternal-cord samples and only chemical concentrations that were above the limit of detection and were not 0. This criterion inevitably eliminates chemicals that partition very strongly to one of the two phases (maternal or cord blood) leaving immeasurable amounts on the other side.

When possible, we selected central tendency R_{CM} values calculated and reported in the paper by the authors. When R_{CM} values were not available, we calculated these values ourselves from the concentrations reported in maternal and cord blood. It should be noted that when we refer chemical concentrations in maternal or cord blood in the manuscript, we refer to serum or plasma. The details of each publication can be accessed by following the link to each study in Supplemental Spreadsheet 1.

2.3 Compiling the Database

Due to the great variability in how studies report the central tendency of R_{CM} (mean, median, geometric mean) and due to the limited number of studies, we collected all reported values and followed a prioritization scheme. We prioritized median over geometric mean and geometric mean over arithmetic mean (median > geometric mean > arithmetic mean) (Fig. 1). Similar approaches have also been used in previous studies.[43] Ideally, we would also want to collect the extremes of the reported values (e.g., min and max). However, given the large variability in how studies report extremes (e.g., min and max, standard deviation, standard error, 25^{th} and 75^{th} percentiles etc.), we focused our efforts on the central tendency as a metric of the ability of a chemical to cross the placenta.

In addition to variability in central tendency, there is also some variability in how studies report chemical concentrations with regard to lipid weight or wet weight. For example, a number of studies measure chemicals that preferentially partition to adipose tissue (e.g., PBDEs and PCBs) and thus they may report chemical measurements per g lipid weight (lipid-adjusted) and/or measurements per g wet weight. Lipid-adjustment for concentrations in blood / serum samples is commonly done by measuring the lipid content of blood /serum and then dividing the measured concentrations in blood by the concentrations of lipids in blood assuming the same volume units. For example, if the measured concentration of a chemical compound x in blood is 4 pg/mL and the concentration of lipids in the sample are 8 mg/mL then the lipid-adjusted concentration of x as follows:

$$[x]_{lipid-adjusted} = \frac{[x]_{\,in\ wet\ weight}}{[lipids]} = \frac{4\frac{pg}{mL}}{8\frac{mg}{mL}} = 0.5\frac{pg}{mg}lipid \quad weight$$

Both concentrations in lipid weight and wet weight can be used to calculate R_{CM}. However, when using lipid-adjusted concentrations, R_{CM} will also be lipid-adjusted due to the differences in the concentrations of lipids between maternal and cord blood.[2,44,45] Lipid-adjusted R_{CM} values are consistently higher than non-adjusted due to the higher lipid content in maternal blood compared to that in cord blood. Previous studies measured and compared the concentration of lipids in maternal and cord blood[2,44,45] and calculated an average ratio (lipids in maternal blood / lipids in cord blood) of 3.18. Considering that all three studies reported similar values after measuring lipid concentrations in different populations, it can be assumed, for modeling purposes, that the difference in the concentration of lipids between maternal and cord samples on average is constant. Thus, one can interpolate between R_{CM} and lipid-adjusted R_{CM} as follows:

$$R_{CM\ lipid\ adjusted} = 3.18 \quad R_{CM}$$

In order to include as many studies as possible in our database, we used the average lipid content ratio (3.18) from three studies[2,44,45] to back-calculate non-lipid adjusted R_{CM} values for the studies that reported only lipid-adjusted R_{CM} values. We converted all values to non-adjusted because calculations of lipid-adjusted values are not meaningful for hydrophilic chemicals or for chemicals that do not partition to either water or lipids, such as PFAS.

Finally, as many compounds may have more than one measurement of R_{CM} in the literature, we calculated the average R_{CM} per compound by averaging across replicates using the arithmetic mean. The resulting dataset with one R_{CM} value per chemical compound was then used to train and evaluate our models. The original database with all measurements and the final file with the averages are provided as Supporting Information in Supplemental Spreadsheet 1 (Also available on GitHub under: https://github.com/dimitriabrahamsson/pfas-maternal-cord).

2.4 Physicochemical descriptors

When curating the chemical structures in the R_{CM} database we tried to adhere to best practices previously reported in the literature.[46,47] For the chemicals in the R_{CM} database, we collected the QSAR-Ready SMILES (simplified molecular-input line-entry system) from the dashboard, which are the de-salted, de-isotoped, stereo-neutral forms of the chemical structures. These QSAR-Ready SMILES were then used to calculate a series of physicochemical descriptors using the cheminformatics package Mordred developed by Moriwaki et al.[48] Mordred can be used to calculate 1,826 2D and 3D physicochemical descriptors, which can then be used as inputs in modeling calculations. These descriptors are primarily low-level physicochemical descriptors covering for example the number of atoms per molecule, number of functional groups, their potential for van der Waals interactions, and their McGowan molecular volume. In that sense, they can cover any molecule with

a known structure that can be represented as a SMILES string. Due to the large size of the datasets, we chose to calculate all the descriptors with Mordred, instead of using experimentally measured descriptors (e.g., for water solubility or hydrophobicity). Including experimental descriptors would require an extensive literature search and an additional compilation strategy for thousands of chemical compounds, which is beyond the scope of the study.

In addition to the Mordred descriptors that were used as inputs for the models, we also collected the RDKit bits[49] (version 2022.3.4; n step = 2) for each chemical in the database, which were used in the statistical analysis to identify the molecular fragments/bits that were significantly associated with R_{CM}. RDKit bits are units of information that describe structural characteristics of a given compound (e.g., C-C-OH). We used a linear regression model to calculate the Pearson r and p-values for every descriptor (Mordred descriptors and RDKit bits) and R_{CM}. To correct for multiple hypothesis testing, we applied the approach of Benjamini-Hochberg using a false discovery rate of 0.05. The process was repeated once for the compounds in the R_{CM} database and once for the compounds on PFASMASTER. We only considered compounds that were within the applicability domain of the model (described below in the applicability domain section).

2.5 Model development, evaluation and application

In deciding which type of algorithm to use for predicting R_{CM}, we evaluated 3 different machine learning algorithms a Support Vector Machine (SVM), a Random Forest (RF), and an Artificial Neural Network (ANN) using the platforms scikit-learn[50] for SVM and RF and Tensorflow[51] for ANN (Fig. 1). All code was written in Python[52] (version 3.9.7) and it is available on GitHub under the following repository: https://github.com/dimitriabrahamsson/pfas-maternal-cord. These types of algorithms are commonly used in quantitative structure activity relationships (QSARs) in the predictions of physicochemical properties or toxicities in environmental and health-related applications.[53–58] The hyperparameters of models were selected using a grid search optimization with a K-fold cross-validation (n=5) using the GridSearchCV package of scikit-learn[50]. The cross-validation was done with an 80/20 split (80% of the dataset in the training set and 20% in the testing set). The hyperparameters were optimized based on mean absolute error (MAE) of the predictions of R_{CM} for the testing set. We then selected the model that showed the lowest MAE for further evaluation. More information on the background of these approaches and on the parametrization for our exercise are presented in the Supporting Information (Text S1). The results of the grid search optimization are presented in Supplemental Spreadsheet 2 and are also available on GitHub (https://github.com/dimitriabrahamsson/pfas-maternal-cord).

2.5.1 Cross-validation—

The best performing model (lowest MAE for the testing set) was then further evaluated using a shuffle-split cross-validation (n=100 times) with an 80/20 split. The performance of the model was evaluated based on the cross-validation mean absolute error (MAE) and on the cross-validation regression coefficient (Q^2) for the testing set. MAE and Q^2 were defined as follows:

$$MAE = \frac{\sum_{i=1}^{n} |y_i - x_i|}{n}$$

Where, y is the prediction and x is the true value for observation i and n is the number of observations.

$$Q^2 = \left(\frac{cov(x, y)}{\sigma_x \sigma_y} \right)^2$$

Where, cov is the covariance, σ_X is the standard deviation of x (true values), σ_Y is the standard deviation of y (predicted values).

2.5.2 Y-randomization to evaluate model predictive power—Finally, the best performing model was evaluated using a y-randomization (n=100 times). Y-randomization is used to test the performance of a model built on the original data by comparing it to a model built on randomly shuffled y-values.[59] This technique allows the evaluation of the model's predictive power relative to randomly generated values. The process was repeated 100 times and every time the model was trained on random y-values. The performance of the model was evaluated by comparing the cross-validation MAE and the cross-validation (Q^2) of the testing set for the real measurements to those of the random values.

2.5.3 Model application and predictions—As an initial application for our model predictions, we used our best performing models to predict the R_{CM} values of 23 PFAS that were included in our database (Supplemental Spreadsheet 1). Similar to the cross-validation step, we repeated the process of predicting R_{CM} for these PFAS 100 times and calculated the averages of these predictions for each chemical. We then compared the predicted values to the experimental R_{CM} values of the 23 PFAS that were present in the database. We then followed the same approach and applied it to all the compounds in PFASMASTER (n=9,252).

2.5.4 Applicability domain—Before defining the applicability domain of our model, it is important to point out that all definitions of an applicability domain for QSARs are inherently subjective. When the true values of a property are unknown, there is no objective way of knowing whether the predictions of the model for that property are truly accurate.[60,61] That being said, defining an applicability domain can help us to make an educated guess of the expected errors of the model by evaluating the similarity of a target compound to compounds in the training set. For the purposes of our study, we defined the applicability domain of our model using the following steps: i) we created a hypothetical compound (compound H) by calculating the average of every physicochemical descriptor (calculated with Mordred) in the training set; ii) we then used a linear regression model to calculate the Pearson R^2 between the physicochemical descriptors of the target compound and those of compound H, which we used as our metric of the applicability domain (AR^2); iii) AR^2 was calculated first for the compounds in the R_{CM} database; and then iv) for all PFAS compounds on PFASMASTER. Calculating AR^2 for the compounds in the R_{CM}

database helps us calculate the lowest boundary of AR^2 beyond which a compound would be considered outside the applicability domain of the model. We defined that lower boundary as the lowest AR^2 observed for the chemicals in the R_{CM} database. That cutoff point was then used to identify which PFAS were outside the applicability domain of the model.

3. Results

3.1 Data collection and description

3.1.1 The R_{CM} values—Our literature search yielded 88 studies and 496 R_{CM} values, of which 323 values were measurements of median R_{CM}, 67 were measurements of geometric mean, 62 were measurements of arithmetic mean, and for 40 values it was unclear from the publication which calculation of central tendency was used. The references for all the studies and all the collected data are presented in Supplemental Spreadsheet 1. After averaging for duplicates, 260 unique chemical structures remained in the database. Some examples of the chemical categories we collected were: PFAS, PCBs, PBDEs, benzophenones, pesticides and pharmaceuticals (Fig. 2; Supplemental Spreadsheet 1). The collected log R_{CM} values ranged from −2.89 to 1.02 and had a median value of −0.25 (Fig. S1). The number of participants in the collected studies ranged from 1 to 1000s with the majority of the studies being in the 10 to 100 participant range. Some of the chemical categories with the highest number of participants were thyroid hormones, plasticizers, phthalates and PFAS. (Fig. S1 and S2)

3.1.2 The R_{CM} values by chemical category—For most chemical categories in the R_{CM} database, the median log R_{CM} was below 0 indicating that most chemicals favorably partition to maternal blood (Fig. 2). However, as for most categories, R_{CM} centered around 0, this also indicates that most chemicals can cross the placenta (Fig. 2) and expose the fetus at concentrations comparable to those in maternal blood. Polychlorinated dibenzofurans and dioxins exhibited the lowest median log R_{CM}. PFAS (n=24) exhibited a median log R_{CM} of about −0.2 indicating slight preference for partitioning to maternal blood. However, 7 PFAS showed preferential partitioning to cord blood (Fig. 2 and Supplemental Spreadsheet 1).

3.1.3 Physicochemical space—Out of 9,252 structures on PFASMASTER, 7,982 compounds had available QSAR-Ready SMILES on EPA's CompTox Chemicals Dashboard. A principal component analysis (PCA) of the physicochemical descriptors showed that the sum of PC1, PC2 and PC3 explained approximately 50% of the variability in the dataset (Fig. S3). We observed a substantial overlap in terms of chemical space between the compounds in our compiled R_{CM} database and the PFAS database. Out of the 7,982 structures, 4,292 structures overlapped in terms of PC1–3 with the R_{CM} database (Fig. S4).

3.2 Model Results

3.2.1 Model evaluation—All three models showed comparable errors for the testing set with an MAE ranging from 0.35 to 0.40 log units (Supplemental Spreadsheet 2). ANN showed the lowest MAE for the testing set with 0.35 log units. The optimized hyperparameters are all presented in Supplemental Spreadsheet 2. Further evaluation of

the model with a shuffle-split cross-validation (n=100) showed a Q^2 of 0.57 for the training set and 0.31 for the testing set, and an MAE of 0.19 log units for the training set and 0.27 for the testing set (Fig. 3).

The y-randomization analysis showed that the model performed substantially worse when trained with random data both in its predictions for the training and testing sets. The Q^2 dropped from 0.57 to 0.36 for the training set and from 0.31 and 0.0017 for the testing set (Fig. 4). This pronounced drop especially for the testing set shows that the model is able to clearly recognize patterns in the training set and apply them to the testing set.

3.2.2 Model application—The model application to the 23 PFAS present in the R_{CM} database (Fig. 5 and Supplemental Spreadsheet 3), showed comparable values of Q^2 and MAE to those of the cross-validation exercise (Fig. 3). The model application of the model on the compounds from the PFASMASTER list produced a distribution with a mean log R_{CM} of −0.031 (Fig. 6), where 4303 compounds had a log $R_{CM} < 0$ indicating preferable partitioning to maternal blood, 3623 compounds had a log $R_{CM} > 0$ indicating preferable partitioning to cord blood, and for 56 compounds the model failed to make predictions (predicted log $R_{CM} = \infty$ or $-\infty$) (Fig. 6).

When examining the applicability domain of our model, we observed that the minimum applicability R^2 value in the R_{CM} database was 0.34 (Fig. 7). Using this value as a cutoff point to determine which compounds from PFASMASTER are expected to be outside the applicability domain of the model, we determined that 7855 compounds were within the applicability domain and 127 compounds are outside the applicability domain (Fig. 7 and Supplemental Spreadsheet 3).

3.3 Significant physicochemical properties and significant molecular fragments

Out of 1825 physicochemical descriptors in total (Mordred), 444 showed a significant association with R_{CM} for the chemicals in the R_{CM} database, and 896 descriptors showed a significant correlation with the predicted R_{CM} values in the PFAS database (after adjusting for multiple hypotheses testing). Among the top physicochemical properties (lowest p-values) that showed significant associations with R_{CM} were water solubility (LogS), van der Waals surface area (e.g., PEOE_VSA6, EState_VSA9), hydrophobicity expressed as the octanol/water equilibrium partition ratio (LogP), molecular weight (MW), and the McGowan molecular volume (VMcGowan) (Supplemental Spreadsheet 3). Increase in water solubility was associated with higher R_{CM}, whereas increase in hydrophobicity, van der Waals surface area, molecular weight and molecular volume were associated with lower R_{CM}.

We observed significant associations between 13 RDKit bits and R_{CM} for the chemicals in the R_{CM} database, and 262 RDKit bits and the predicted R_{CM} values in the PFAS database (after adjusting for multiple hypotheses testing). Our analysis of the RDKit bits and their relationship with R_{CM} of the chemicals in the maternal-cord database showed that chemicals that show higher R_{CM} (preferential partitioning in cord blood) contain polar functional groups such as C-O-O, C-N and C-C-N, while chemicals that show lower R_{CM} (preferential partitioning in maternal blood) contain non-polar functional groups, such as

C-C-Cl, C-Cl and C-C-S (Fig. 8). This observation is also reflected in the predicted R_{CM} values for PFAS. PFAS that have a higher R_{CM} (preferential partitioning in cord blood) contain polar functional groups such as O-H, C-O-O-H, C-O-O, and F-C-O, while PFAS that show a lower R_{CM} contain more non-polar functional groups such as, C(aromatic)-F, C(aromatic)-C-F, and C-S (Fig. 8). However, for some PFAS, there are some polar groups also in the compounds that showed a higher R_{CM}. Some examples are N-C (triple bond) and O-S-O (double bonds).

4. Discussion

Our study generated a large database of R_{CM} values that can be used to further examine the transplacental transfer of small molecules and their partitioning between maternal and cord blood (Supplemental Spreadsheet 1). To our knowledge, our compiled R_{CM} database is the largest publicly available database of R_{CM} measurements from matched maternal and cord blood samples. The predictions of R_{CM} for the compounds in PFASMASTER centered around 0 with the distribution almost split in half between compounds that partition favorably to maternal blood and compounds that partition favorably to cord blood. For 3623 compounds in PFASMASTER log R_{CM} was higher than 0 indicating preferable partitioning to cord blood. This observation has important implications for human health risks as compounds that cross the placenta and partition to cord blood could pose a greater risk to the developing fetus. To put things into a broader perspective, we compared the R_{CM} values calculated for PFAS to two antidepressant pharmaceuticals, Fluvoxamine and Fluoxetine, that are often prescribed during pregnancy and have been shown to have very limited side effects to the fetus.[62–64] As Fluvoxamine and Fluoxetine exhibit log R_{CM} values of −3 and −2 respectively (Fig. 6 and Supplementary Spreadsheet 1), their limited impact on fetal health can be in part attributed to their inability to efficiently cross the placenta. With a log R_{CM} value of −3, for example, the concentration of Fluvoxamine in cord blood is expected to be 1000 times lower than the concentration in maternal blood. Compared to these compounds, most of the PFAS we tested in this study are expected to crossing the placenta and expose the fetus.

Our analysis of the chemical fragments (RDKit bits) and their associations with R_{CM} showed that chemicals with polar groups, such as CN and COOH, are more likely to cross the placental barrier and partition to cord blood (Fig. 8). This is reflected both in the R_{CM} database with the experimental values and in the PFAS database with the predicted data (Fig. 8). However, in the case of PFAS, there were also some polar groups, such as CN (triple bond) and SO2 (double bonds) in the chemicals that are predicted to partition preferentially to maternal blood. This finding likely indicates that even though polarity can explain some of the variability in R_{CM}, the process of transplacental transfer is a more complex process where other factors also play an important role. These observations are in agreement with our conclusions from the analysis of the physicochemical properties (Mordred descriptors) and their associations with R_{CM} which showed that water solubility (logS), van der Waals surface area, hydrophobicity (logP), molecular weight (MW) and the McGowan molecular volume (VMcGowan) were among the most significant parameters (lowest p-values) associated with R_{CM} in the R_{CM} database. These parameters were also among the most significant parameters in the R_{CM} predictions for the PFAS database.

Compounds with higher water solubility, lower van der Waals surface area and smaller weight and molecular volumes were more likely to partition to cord blood (Supplemental Spreadsheet 3). Some examples of PFAS that are expected to cross the placenta and preferentially partition to cord blood are shown in Fig. 9.

Comparing the findings in this report to previous studies that examined how physicochemical properties influence the transplacental transfer of small molecules, we came across some interesting similarities but also some important differences. Our findings are in agreement with the study of Jeong et al.[65] who observed that an increase in the numbers of chlorines, molecular weight and hydrophobicity were associated with a less efficient transplacental transfer and consequently lower concentrations in the cord blood.

Li et al. [66] observed that the partitioning of small molecules between maternal and cord blood can be described by a set of physicochemical properties which explained 66.5% of the variability in R_{CM} of their dataset (n = 51). While this is similar to our findings of the PCA analysis (Fig. S3 and S4), our studies differ on how the physicochemical properties affect the partitioning of chemicals between maternal and cord blood. Li et al. [66] suggested that the number of heavier atoms, the number of halogens, the mean atomic Van der Waals volume, and the McGowan molecular volume were positively associated with R_{CM}, while the number of oxygen atoms and the topological polar surface area were negatively associated with R_{CM}. In our study, however, we observed that an increase in the number of halogens and the McGowan molecular volume was associated with a decrease in R_{CM} while an increase in polarity and water solubility was associated with an increase in R_{CM}. It is important to note that the dataset we used in our study is about 5 times larger (n = 260) than that of Li et al. (n = 51) and perhaps the apparent disagreement is reflective of a difference in data size and structural variability.

It is important to also note that R_{CM} is known to be influenced by physiological parameters such as gestational age[4] and gestational diabetes[3]. These observations are in good agreement with the conventional thermodynamic understanding that the behavior and fate of chemicals in the environment is controlled by their physicochemical properties and the properties of their environment. Li et al.[4] noted that overall transplacental transfer was higher in full term birth compared to preterm birth (overall increase in R_{CM}: 0.12 – 0.54), and Eryasa et al.[3] observed that overall transplacental transfer was higher in mothers with gestational diabetes[3] (overall increase in R_{CM}: 0.13 – 0.44). This further indicates a limitation, and future need, in this research area. Our analysis focused on central tendency values for R_{CM}. However, factors such as preexisting conditions, in addition to other biological factors, such as genetics and age are likely to influence the variability of the maternal cord partitioning of chemicals as they change the properties of the environment where chemicals partition in. Evaluating the upper or lower confidence limits of R_{CM} is a challenging task due to the great variability on how studies report the upper and lower limits (e.g., min and max, standard deviation, standard error, 25th and 75th percentiles). It is, however, an important aspect that needs to be addressed in future studies to better characterize the variability in R_{CM}.

Finally, we should note that as the collected R_{CM} values from the literature are composed of arithmetic means, geometric means and medians, there may be some noise associated

with the modeled endpoint. However, we should also point out that 65% of the values are medians leaving 35% for geometric means and arithmetic means. Thus, we would expect that the effect of that noise would likely be moderate.

To our knowledge, this is the first study to predict R_{CM} values for over 7,000 PFAS compounds. Approximately, half of the PFAS in the database show a predicted preferential partitioning to cord blood, which is concerning considering the toxicity potential of many PFAS that has been demonstrated in toxicological and epidemiological studies. Our findings underscore the importance of regulation of PFAS and the need for safer alternatives. We should note that while our study focuses on PFAS as a case study, our model is applicable to other chemical categories also as long as their molecular structures are known, and they are within the applicability domain of the model. EPA's CompTox chemicals dashboard lists about 1,000,000 chemicals that are expected to be of environmental importance. For the vast majority of these chemicals, there is very little information on their toxicity or their potential to expose the fetus during pregnancy. Some categories that would be of interest for future studies are plasticizers, pesticides and chemicals used in personal care products.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

We want to thank Nisha Sipes (U.S. EPA), Matthew MacLeod (Stockholm University), and Ian Cousins (Stockholm University) for their thoughtful comments and immensely valuable suggestions.

Funding

This study was funded by the Office of Environmental Health Hazard Assessment (OEHHA) of the California Environmental Protection Agency (CalEPA) and by the National Institutes of Health / National Institute of Environmental Health Sciences (NIH/NIEHS) (grant numbers K99ES032892, P30-ES030284, P01ES022841, R01ES027051). The views expressed in this manuscript are those of the authors and do not necessarily represent the views and policies of OEHHA.

Data availability statement

The manuscript is accompanied by a supporting information document file and 3 supplemental spreadsheets with the all the underlying data (Supplemental Spreadsheet 1 – Database, Supplemental Spreadsheet 2 – Grid Search, Supplemental Spreadsheet 3 – Modeling Results). All code, supplemental documents and spreadsheets are available on GitHub (https://github.com/dimitriabrahamsson/pfas-maternal-cord).

References

(1). Huppertz B The Anatomy of the Normal Placenta. Journal of Clinical Pathology 2008, 61 (12), 1296–1302. 10.1136/jcp.2008.055277. [PubMed: 18755720]

(2). Needham LL; Grandjean P; Heinzow B; Jørgensen PJ; Nielsen F; Patterson DG; et al. Partition of Environmental Chemicals between Maternal and Fetal Blood and Tissues. Environmental science & technology 2011, 45 (3), 1121–1126. 10.1021/es1019614. [PubMed: 21166449]

(3). Eryasa B; Grandjean P; Nielsen F; Valvi D; Zmirou-Navier D; Sunderland E; et al. Physico-Chemical Properties and Gestational Diabetes Predict Transplacental Transfer and

Partitioning of Perfluoroalkyl Substances. Environment International 2019, 130, 104874. 10.1016/j.envint.2019.05.068. [PubMed: 31200157]

(4). Li J; Cai D; Chu C; Li Q; Zhou Y; Hu L; et al. Transplacental Transfer of Per- and Polyfluoroalkyl Substances (PFASs): Differences between Preterm and Full-Term Deliveries and Associations with Placental Transporter MRNA Expression. Environ. Sci. Technol 2020, 54 (8), 5062–5070. 10.1021/acs.est.0c00829. [PubMed: 32208722]

(5). Chen A; Park J-S; Linderholm L; Rhee A; Petreas M; DeFranco EA; et al. Hydroxylated Polybrominated Diphenyl Ethers in Paired Maternal and Cord Sera. Environ. Sci. Technol 2013, 47 (8), 3902–3908. 10.1021/es3046839. [PubMed: 23506475]

(6). Li M; Zeng X-W; Qian Z. (Min); Vaughn MG; Sauvé S; Paul G; et al. Isomers of Perfluorooctanesulfonate (PFOS) in Cord Serum and Birth Outcomes in China: Guangzhou Birth Cohort Study. Environment International 2017, 102, 1–8. 10.1016/j.envint.2017.03.006. [PubMed: 28297681]

(7). Ferguson KK; Rosen EM; Rosario Z; Feric Z; Calafat AM; McElrath TF; et al. Environmental Phthalate Exposure and Preterm Birth in the PROTECT Birth Cohort. Environ Int 2019, 132, 105099. 10.1016/j.envint.2019.105099. [PubMed: 31430608]

(8). Syme MR; Paxton JW; Keelan JA Drug Transfer and Metabolism by the Human Placenta. Clin Pharmacokinet 2004, 43 (8), 487–514. 10.2165/00003088-200443080-00001. [PubMed: 15170365]

(9). Wang A; Padula A; Sirota M; Woodruff TJ Environmental Influences on Reproductive Health: The Importance of Chemical Exposures. Fertil. Steril 2016, 106 (4), 905–929. 10.1016/j.fertnstert.2016.07.1076. [PubMed: 27513554]

(10). Domingo JL; Nadal M Human Exposure to Per- and Polyfluoroalkyl Substances (PFAS) through Drinking Water: A Review of the Recent Scientific Literature. Environmental Research 2019, 177, 108648. 10.1016/j.envres.2019.108648. [PubMed: 31421451]

(11). Sunderland EM; Hu XC; Dassuncao C; Tokranov AK; Wagner CC; Allen JG A Review of the Pathways of Human Exposure to Poly- and Perfluoroalkyl Substances (PFASs) and Present Understanding of Health Effects. Journal of Exposure Science & Environmental Epidemiology 2019, 29 (2), 131–147. 10.1038/s41370-018-0094-1. [PubMed: 30470793]

(12). U.S. Environmental Protection Agency. Chemistry Dashboard. https://comptox.epa.gov/dashboard/ (accessed 2021-03-09).

(13). Williams AJ; Grulke CM; Edwards J; McEachran AD; Mansouri K; Baker NC; et al. The CompTox Chemistry Dashboard: A Community Data Resource for Environmental Chemistry. J Cheminform 2017, 9 (1), 61. 10.1186/s13321-017-0247-6. [PubMed: 29185060]

(14). Highly fluorinated substances. https://www.kemi.se/en/chemical-substances-and-materials/highly-fluorinated-substances (accessed 2020-12-08).

(15). Giesy JP; Kannan K Global Distribution of Perfluorooctane Sulfonate in Wildlife. Environ. Sci. Technol 2001, 35 (7), 1339–1342. 10.1021/es001834k. [PubMed: 11348064]

(16). Martin JW; Mabury SA; Solomon KR; Muir DCG Bioconcentration and Tissue Distribution of Perfluorinated Acids in Rainbow Trout (Oncorhynchus Mykiss). Environmental Toxicology and Chemistry 2003, 22 (1), 196–204. 10.1002/etc.5620220126. [PubMed: 12503765]

(17). Gaballah S; Swank A; Sobus JR; Howey XM; Schmid J; Catron T; et al. Evaluation of Developmental Toxicity, Developmental Neurotoxicity, and Tissue Dose in Zebrafish Exposed to GenX and Other PFAS. Environmental Health Perspectives 128 (4), 047005. 10.1289/EHP5843.

(18). Ahrens L; Bundschuh M Fate and Effects of Poly- and Perfluoroalkyl Substances in the Aquatic Environment: A Review. Environmental Toxicology and Chemistry 2014, 33 (9), 1921–1929. 10.1002/etc.2663. [PubMed: 24924660]

(19). Fenton SE; Ducatman A; Boobis A; DeWitt JC; Lau C; Ng C; et al. Per- and Polyfluoroalkyl Substance Toxicity and Human Health Review: Current State of Knowledge and Strategies for Informing Future Research. Environmental Toxicology and Chemistry 2021, 40 (3), 606–630. 10.1002/etc.4890. [PubMed: 33017053]

(20). Grandjean P; Andersen EW; Budtz-Jørgensen E; Nielsen F; Mølbak K; Weihe P; et al. Serum Vaccine Antibody Concentrations in Children Exposed to Perfluorinated Compounds. JAMA 2012, 307 (4), 391–397. 10.1001/jama.2011.2034. [PubMed: 22274686]

(21). Grandjean P; Heilmann C; Weihe P; Nielsen F; Mogensen UB; Budtz-Jørgensen E Serum Vaccine Antibody Concentrations in Adolescents Exposed to Perfluorinated Compounds. Environ Health Perspect 2017, 125 (7), 077018. 10.1289/EHP275. [PubMed: 28749778]

(22). C8 Science Panel Website. http://www.c8sciencepanel.org/prob_link.html (accessed 2020-12-08).

(23). Yamaguchi M; Arisawa K; Uemura H; Katsuura-Kamano S; Takami H; Sawachika F; et al. Consumption of Seafood, Serum Liver Enzymes, and Blood Levels of PFOS and PFOA in the Japanese Population. Journal of Occupational Health 2013, 55 (3), 184–194. 10.1539/joh.12-0264-OA. [PubMed: 23574777]

(24). Massoud O; Charlton M Nonalcoholic Fatty Liver Disease/Nonalcoholic Steatohepatitis and Hepatocellular Carcinoma. Clinics in Liver Disease 2018, 22 (1), 201–211. 10.1016/j.cld.2017.08.014. [PubMed: 29128057]

(25). Šabović I; Cosci I; De Toni L; Ferramosca A; Stornaiuolo M; Di Nisio A; et al. Perfluoro-Octanoic Acid Impairs Sperm Motility through the Alteration of Plasma Membrane. J Endocrinol Invest 2020, 43 (5), 641–652. 10.1007/s40618-019-01152-0. [PubMed: 31776969]

(26). Lam J; Koustas E; Sutton P; Johnson PI; Atchley DS; Sen S; et al. The Navigation Guide - Evidence-Based Medicine Meets Environmental Health: Integration of Animal and Human Evidence for PFOA Effects on Fetal Growth. Environ Health Perspect 2014, 122 (10), 1040–1051. 10.1289/ehp.1307923. [PubMed: 24968389]

(27). Chen F; Yin S; Kelly BC; Liu W Isomer-Specific Transplacental Transfer of Perfluoroalkyl Acids: Results from a Survey of Paired Maternal, Cord Sera, and Placentas. Environ. Sci. Technol 2017, 51 (10), 5756–5763. 10.1021/acs.est.7b00268. [PubMed: 28434222]

(28). Morello-Frosch R; Cushing LJ; Jesdale BM; Schwartz JM; Guo W; Guo T; et al. Environmental Chemicals in an Urban Population of Pregnant Women and Their Newborns from San Francisco. Environmental Science & Technology 2016, 50 (22), 12464–12472. 10.1021/acs.est.6b03492. [PubMed: 27700069]

(29). Gao K; Zhuang T; Liu X; Fu J; Zhang J; Fu J; et al. Prenatal Exposure to Per- and Polyfluoroalkyl Substances (PFASs) and Association between the Placental Transfer Efficiencies and Dissociation Constant of Serum Proteins–PFAS Complexes, Environmental Science & Technology 2019, 53 (11), 6529–6538. 10.1021/acs.est.9b00715. [PubMed: 31099564]

(30). Starling AP; Adgate JL; Hamman RF; Kechris K; Calafat AM; Ye X; et al. Perfluoroalkyl Substances during Pregnancy and Offspring Weight and Adiposity at Birth: Examining Mediation by Maternal Fasting Glucose in the Healthy Start Study. Environmental Health Perspectives 125 (6), 067016. 10.1289/EHP641.

(31). Spratlen MJ; Perera FP; Lederman SA; Robinson M; Kannan K; Herbstman J; et al. The Association Between Perfluoroalkyl Substances and Lipids in Cord Blood. J Clin Endocrinol Metab 2020, 105 (1). 10.1210/clinem/dgz024.

(32). Chen Q; Zhang X; Zhao Y; Lu W; Wu J; Zhao S; et al. Prenatal Exposure to Perfluorobutanesulfonic Acid and Childhood Adiposity: A Prospective Birth Cohort Study in Shanghai, China. Chemosphere 2019, 226, 17–23. 10.1016/j.chemosphere.2019.03.095. [PubMed: 30908964]

(33). Høyer BB; Ramlau-Hansen CH; Obel C; Pedersen HS; Hernik A; Ogniev V; et al. Pregnancy Serum Concentrations of Perfluorinated Alkyl Substances and Offspring Behaviour and Motor Development at Age 5–9 Years – a Prospective Study. Environ Health 2015, 14 (1), 2. 10.1186/1476-069X-14-2. [PubMed: 25567242]

(34). Cahill TM; Cousins I; Mackay D Development and Application of a Generalized Physiologically Based Pharmacokinetic Model for Multiple Environmental Contaminants. Environmental Toxicology and Chemistry 2003, 22 (1), 26–34. 10.1002/etc.5620220104. [PubMed: 12503743]

(35). Verner M-A; Loccisano AE; Morken N-H; Yoon M; Wu H; McDougall R; et al. Associations of Perfluoroalkyl Substances (PFAS) with Lower Birth Weight: An Evaluation of Potential Confounding by Glomerular Filtration Rate Using a Physiologically Based Pharmacokinetic Model (PBPK). Environmental Health Perspectives 2015, 123 (12), 1317–1324. 10.1289/ehp.1408837. [PubMed: 26008903]

(36). Forns J; Verner M-A; Iszatt N; Nowack N; Bach CC; Vrijheid M; et al. Early Life Exposure to Perfluoroalkyl Substances (PFAS) and ADHD: A Meta-Analysis of Nine European Population-Based Studies. Environmental Health Perspectives 128 (5), 057002. 10.1289/EHP5444.

(37). Verner M-A; Ngueta G; Jensen ET; Fromme H; Völkel W; Nygaard UC; et al. A Simple Pharmacokinetic Model of Prenatal and Postnatal Exposure to Perfluoroalkyl Substances (PFASs). Environ. Sci. Technol 2016, 50 (2), 978–986. 10.1021/acs.est.5b04399. [PubMed: 26691063]

(38). Mamsen LS; Björvang RD; Mucs D; Vinnars M-T; Papadogiannakis N; Lindh CH; et al. Concentrations of Perfluoroalkyl Substances (PFASs) in Human Embryonic and Fetal Organs from First, Second, and Third Trimester Pregnancies. Environment International 2019, 124, 482–492. 10.1016/j.envint.2019.01.010. [PubMed: 30684806]

(39). Mamsen LS; Jönsson BAG; Lindh CH; Olesen RH; Larsen A; Ernst E; et al. Concentration of Perfluorinated Compounds and Cotinine in Human Foetal Organs, Placenta, and Maternal Plasma. Science of The Total Environment 2017, 596–597, 97–105. 10.1016/j.scitotenv.2017.04.058.

(40). Pérez F; Nadal M; Navarro-Ortega A; Fàbrega F; Domingo JL; Barceló D; et al. Accumulation of Perfluoroalkyl Substances in Human Tissues. Environment International 2013, 59, 354–362. 10.1016/j.envint.2013.06.004. [PubMed: 23892228]

(41). Takaku T; Nagahori H; Sogame Y; Takagi T Quantitative Structure–Activity Relationship Model for the Fetal–Maternal Blood Concentration Ratio of Chemicals in Humans. Biological and Pharmaceutical Bulletin 2015, 38 (6), 930–934. 10.1248/bpb.b14-00883. [PubMed: 26027836]

(42). Lancz K; Murínová Ľ; Patayová H; Drobná B; Wimmerová S; Šovčíková E; et al. Ratio of Cord to Maternal Serum PCB Concentrations in Relation to Their Congener-Specific Physicochemical Properties. International Journal of Hygiene and Environmental Health 2015, 218 (1), 91–98. 10.1016/j.ijheh.2014.08.003. [PubMed: 25277764]

(43). Aylward LL; Hays SM; Kirman CR; Marchitti SA; Kenneke JF; English C; et al. Relationships of Chemical Concentrations in Maternal and Cord Blood: A Review of Available Data. Journal of toxicology and environmental health. Part B, Critical reviews 2014, 17 (3), 175–203. 10.1080/10937404.2014.884956. [PubMed: 24749481]

(44). Nakamura T; Nakai K; Matsumura T; Suzuki S; Saito Y; Satoh H Determination of Dioxins and Polychlorinated Biphenyls in Breast Milk, Maternal Blood and Cord Blood from Residents of Tohoku, Japan. Science of The Total Environment 2008, 394 (1), 39–51. 10.1016/j.scitotenv.2008.01.012. [PubMed: 18291442]

(45). Park J-S; Bergman Å; Linderholm L; Athanasiadou M; Kocan A; Petrik J; et al. Placental Transfer of Polychlorinated Biphenyls, Their Hydroxylated Metabolites and Pentachlorophenol in Pregnant Women from Eastern Slovakia. Chemosphere 2008, 70 (9), 1676–1684. 10.1016/j.chemosphere.2007.07.049. [PubMed: 17764717]

(46). Fourches D; Muratov E; Tropsha A Trust, But Verify: On the Importance of Chemical Structure Curation in Cheminformatics and QSAR Modeling Research. J. Chem. Inf. Model 2010, 50 (7), 1189–1204. 10.1021/ci100176x. [PubMed: 20572635]

(47). Fourches D; Muratov E; Tropsha A Trust, but Verify II: A Practical Guide to Chemogenomics Data Curation. J. Chem. Inf. Model 2016, 56 (7), 1243–1252. 10.1021/acs.jcim.6b00129. [PubMed: 27280890]

(48). Moriwaki H; Tian Y-S; Kawashita N; Takagi T Mordred: A Molecular Descriptor Calculator. Journal of Cheminformatics 2018, 10 (1), 4. 10.1186/s13321-018-0258-y. [PubMed: 29411163]

(49). Landrum G RDKit: Using the New Fingerprint Bit Rendering Code. RDKit, 2018.

(50). scikit-learn: machine learning in Python — scikit-learn 0.22.1 documentation. https://scikit-learn.org/stable/ (accessed 2020-02-20).

(51). TensorFlow. TensorFlow. https://www.tensorflow.org/ (accessed 2020-02-17).

(52). Welcome to Python.org. Python.org. https://www.python.org/ (accessed 2020-02-20).

(53). Baskin II; Palyulin VA; Zefirov NS Neural Networks in Building QSAR Models. In Artificial Neural Networks: Methods and Applications; Livingstone DJ, Ed.; Methods in Molecular BiologyTM; Humana Press: Totowa, NJ, 2009; pp 133–154. 10.1007/978-1-60327-101-1_8.

(54). Dearden JC; Rowe PH Use of Artificial Neural Networks in the QSAR Prediction of Physicochemical Properties and Toxicities for REACH Legislation. In Artificial Neural Networks; Cartwright H, Ed.; Methods in Molecular Biology; Springer: New York, NY, 2015; pp 65–88. 10.1007/978-1-4939-2239-0_5.

(55). Polishchuk PG; Muratov EN; Artemenko AG; Kolumbin OG; Muratov NN; Kuz'min VE Application of Random Forest Approach to QSAR Prediction of Aquatic Toxicity. J. Chem. Inf. Model 2009, 49 (11), 2481–2488. 10.1021/ci900203n. [PubMed: 19860412]

(56). Svetnik V; Liaw A; Tong C; Culberson JC; Sheridan RP; Feuston BP Random Forest: A Classification and Regression Tool for Compound Classification and QSAR Modeling. J. Chem. Inf. Comput. Sci 2003, 43 (6), 1947–1958. 10.1021/ci034160g. [PubMed: 14632445]

(57). Darnag R; Mostapha Mazouz EL; Schmitzer A; Villemin D; Jarid A; Cherqaoui D Support Vector Machines: Development of QSAR Models for Predicting Anti-HIV-1 Activity of TIBO Derivatives. European Journal of Medicinal Chemistry 2010, 45 (4), 1590–1597. 10.1016/j.ejmech.2010.01.002. [PubMed: 20110136]

(58). Maltarollo VG; Kronenberger T; Espinoza GZ; Oliveira PR; Honorio KM Advances with Support Vector Machines for Novel Drug Discovery. Expert Opinion on Drug Discovery 2019, 14 (1), 23–33. 10.1080/17460441.2019.1549033. [PubMed: 30488731]

(59). Rücker C; Rücker G; Meringer M Y-Randomization and Its Variants in QSPR/QSAR. J. Chem. Inf. Model 2007, 47 (6), 2345–2357. 10.1021/ci700157b. [PubMed: 17880194]

(60). Aniceto N; Freitas AA; Bender A; Ghafourian T A Novel Applicability Domain Technique for Mapping Predictive Reliability across the Chemical Space of a QSAR: Reliability-Density Neighbourhood. Journal of Cheminformatics 2016, 8 (1), 69. 10.1186/s13321-016-0182-y.

(61). Netzeva TI; Worth AP; Aldenberg T; Benigni R; Cronin MTD; Gramatica P; et al. Current Status of Methods for Defining the Applicability Domain of (Quantitative) Structure-Activity Relationships: The Report and Recommendations of ECVAM Workshop 521,2. Altern Lab Anim 2005, 33 (2), 155–173. 10.1177/026119290503300209. [PubMed: 16180989]

(62). Gentile S The Safety of Newer Antidepressants in Pregnancy and Breastfeeding. Drug-Safety 2005, 28 (2), 137–152. 10.2165/00002018-200528020-00005. [PubMed: 15691224]

(63). Hallberg P; Sjöblom V The Use of Selective Serotonin Reuptake Inhibitors During Pregnancy and Breast-Feeding: A Review and Clinical Aspects. Journal of Clinical Psychopharmacology 2005, 25 (1), 59–73. 10.1097/01.jcp.0000150228.61501.e4. [PubMed: 15643101]

(64). Patil AS; Kuller JA; Rhee EHJ Antidepressants in Pregnancy: A Review of Commonly Prescribed Medications. Obstetrical & Gynecological Survey 2011, 66 (12), 777–787. 10.1097/OGX.0b013e31823e0cbf. [PubMed: 22192462]

(65). Jeong Y; Lee S; Kim S; Park J; Kim H-J; Choi G; et al. Placental Transfer of Persistent Organic Pollutants and Feasibility Using the Placenta as a Non-Invasive Biomonitoring Matrix. Science of The Total Environment 2018, 612, 1498–1505. 10.1016/j.scitotenv.2017.07.054. [PubMed: 28910963]

(66). Li J; Sun X; Xu J; Tan H; Zeng EY; Chen D Transplacental Transfer of Environmental Chemicals: Roles of Molecular Descriptors and Placental Transporters. Environ. Sci. Technol 2021, 55 (1), 519–528. 10.1021/acs.est.0c06778. [PubMed: 33295769]

Impact statement

Understanding the behavior of chemicals in the human body during pregnancy is critical in preventing harmful exposures during critical periods of development. Many chemicals can cross the placenta and expose the fetus, however, the mechanism by which this transport occurs is not well understood. In our study, we developed a machine learning model that describes the transplacental transfer of chemicals as a function of their physicochemical properties. The model was then used to make predictions for a set of about 9,000 per- polyfluorinated alkyl substances that are listed on EPA's CompTox Chemicals Dashboard.

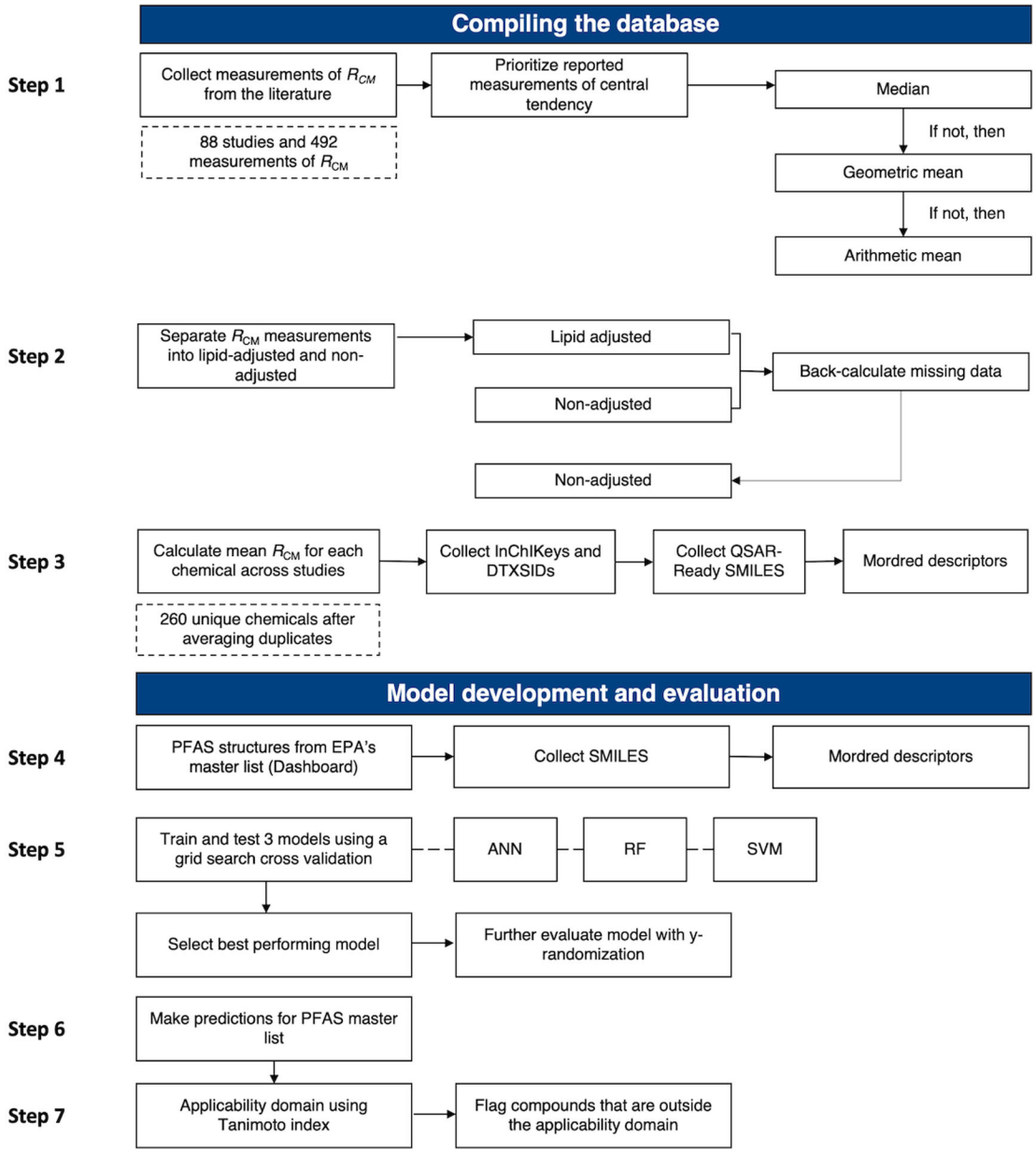

Figure 1:

Workflow for compiling the R_{CM} database and for training and testing the models. The database workflow consisted of three main steps: 1) collection of R_{CM} values from the literature, 2) back-calculating missing values from lipid-adjusted values and completing non-lipid adjusted data and 3) collection of chemical identifiers, SMILES and physicochemical descriptors (Mordred descriptors). The modeling workflow consisted of data processing, model development and model application in predicting R_{CM} for PFAS in the PFAS master list from EPA's CompTox Chemicals Dashboard.

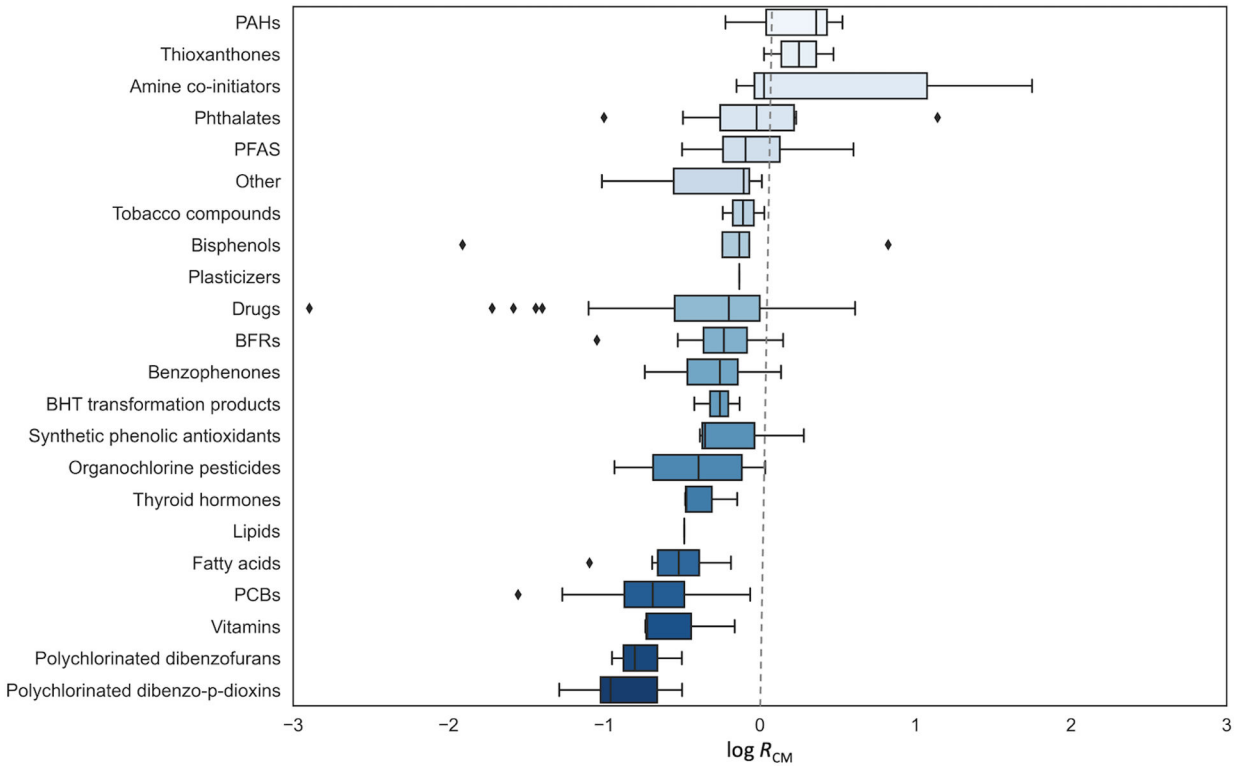

Figure 2:

Log non-lipid adjusted R_{CM} of the chemicals in the compiled database presented by chemical category and sorted by median R_{CM} from the highest to the lowest. The gray dashed line shows the 0 value on the x-axis representing the point at which the concentration of the chemicals in maternal blood is the same as the concentration in cord blood. The underlying data are available in Supplemental Spreadsheet 1.

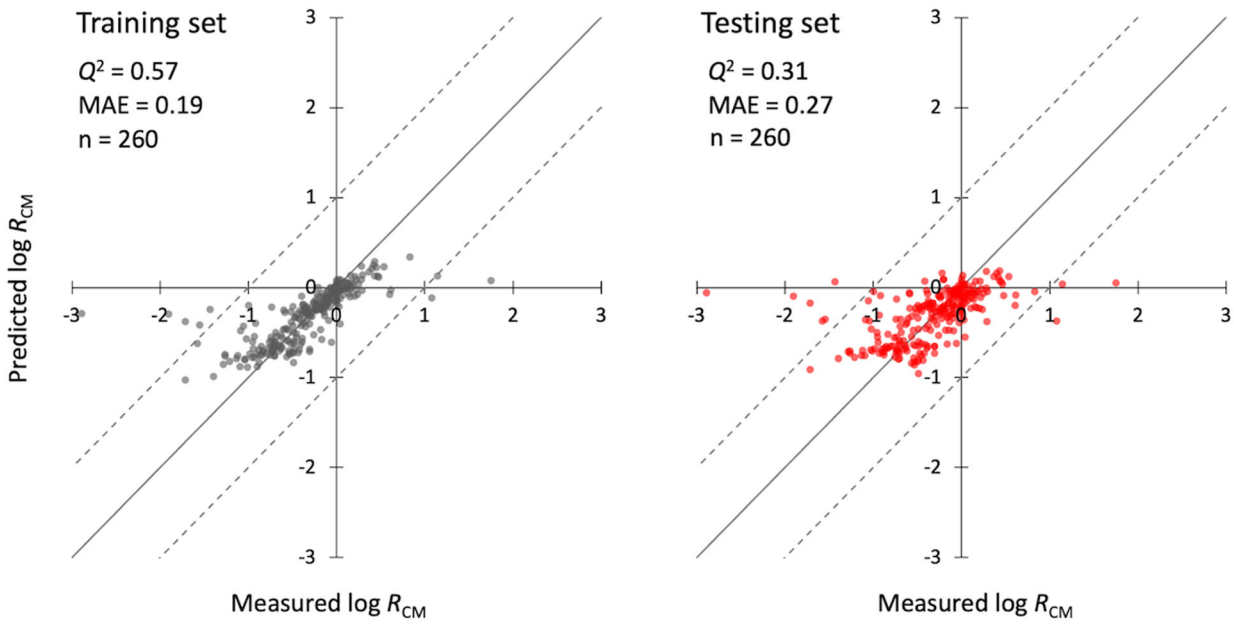

Figure 3:

Cross validation using a shuffle-split method (n=100 times) for dividing the data set into training and testing with an 80:20 split. The results of the cross-validation are presented as true vs predicted values for the best performing ANN model. The solid diagonal line is the 1-to-1 agreement line and the dashed lines show the ± 1 log unit deviation from the 1-to-1 agreement line. Each dot shows the average of predictions for 1 chemical from 100 iterations. The n is the same for both the training and the testing set as after 100 iterations all chemicals in the database had at least one chance to be in the training and in the testing sets.

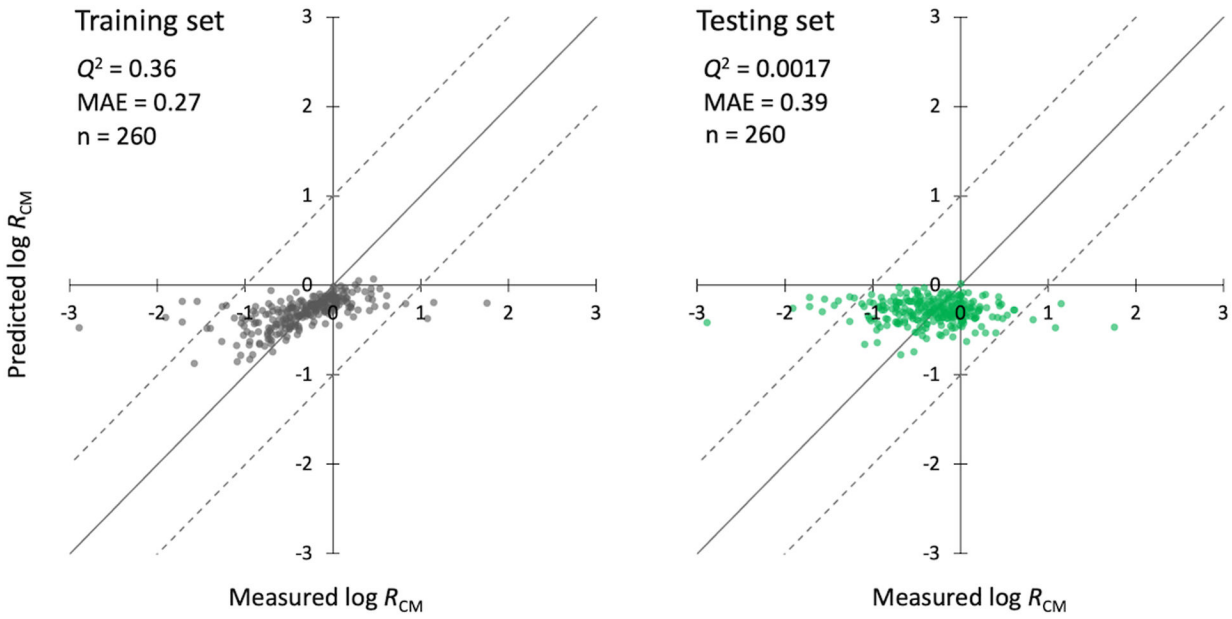

Figure 4:

Y-randomization analysis following a cross-validation using a shuffle-split method (n=100 times) for dividing the data set into training and testing with an 80:20 split. The results of the cross-validation are presented as measured vs predicted values for the best performing ANN model. The solid diagonal line is the 1-to-1 agreement line and the dashed lines show the ± 1 log unit deviation from the 1-to-1 line. Each dot shows the average of predictions for 1 chemical from 100 iterations. The n is the same for both the training and the testing set as after 100 iterations all chemicals in the database had at least one chance to be in the training and in the testing sets.

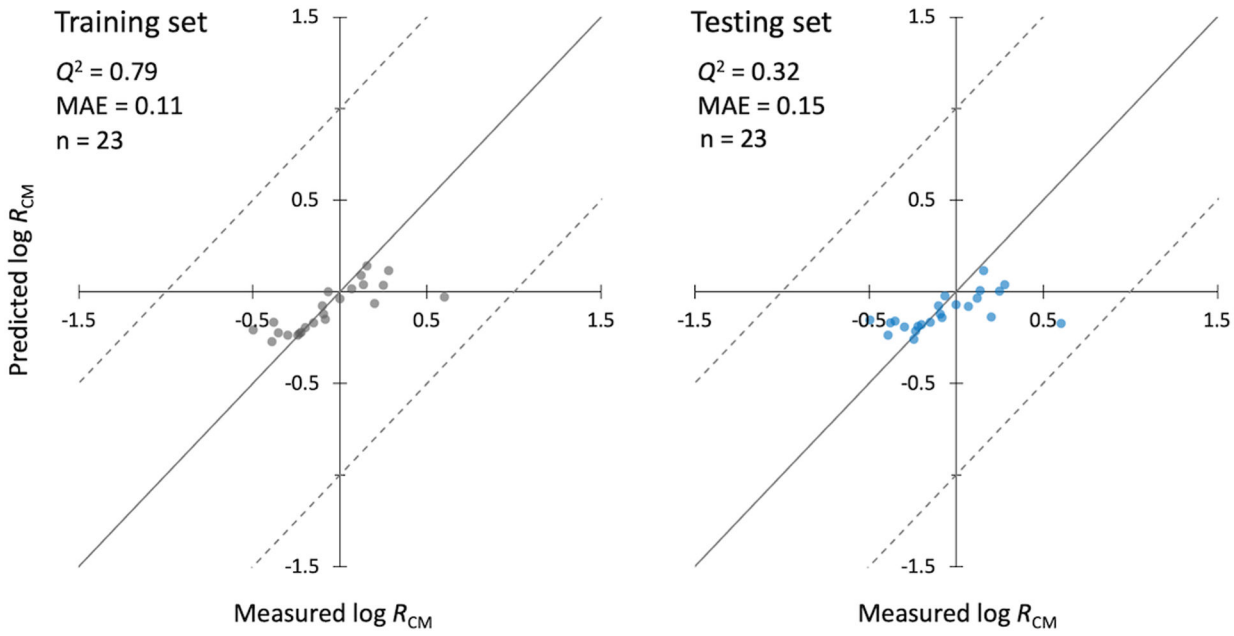

Figure 5:

Predictions of log R_{CM} for the 23 PFAS from the maternal-cord database when the 23 compounds were included in the training set and in when they were included in the testing test. Each dot represents the average of 100 iterations following a shuffle-split cross-validation. The solid diagonal line is the 1-to-1 agreement line and the dashed lines show the ± 0.1 log unit deviation from the 1-to-1 agreement line. The n is the same for both the training and the testing set as after 100 iterations all chemicals in the database had at least one chance to be in the training and in the testing sets.

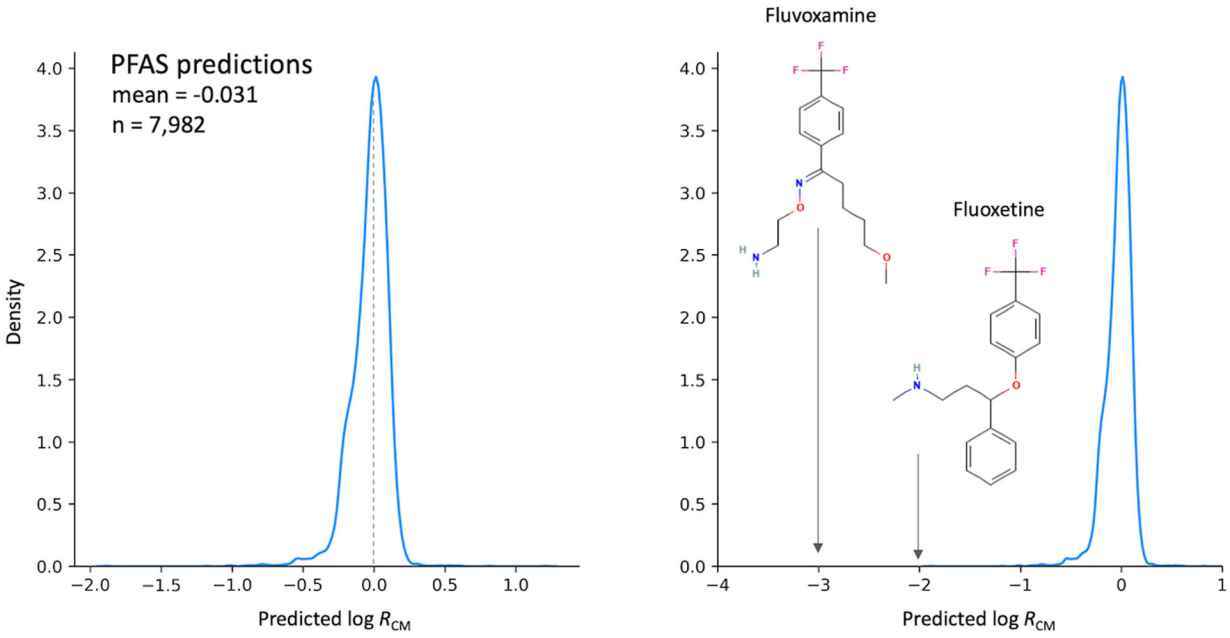

Figure 6:

Model predictions for PFAS in the PFASMASTER list shown as a kernel density estimation plot. For comparison purposes, we also show the reported R_{CM} of Fluvoxamine and Fluoxetine, which are two anti-depressant pharmaceuticals that are commonly prescribed during pregnancy and are considered safe. The dashed vertical line shows 0 for the log R_{CM} which corresponds to a ratio of 1 in linear space and represents the point at which the concentration of a chemical in the maternal blood is equal to the concentration in cord blood.

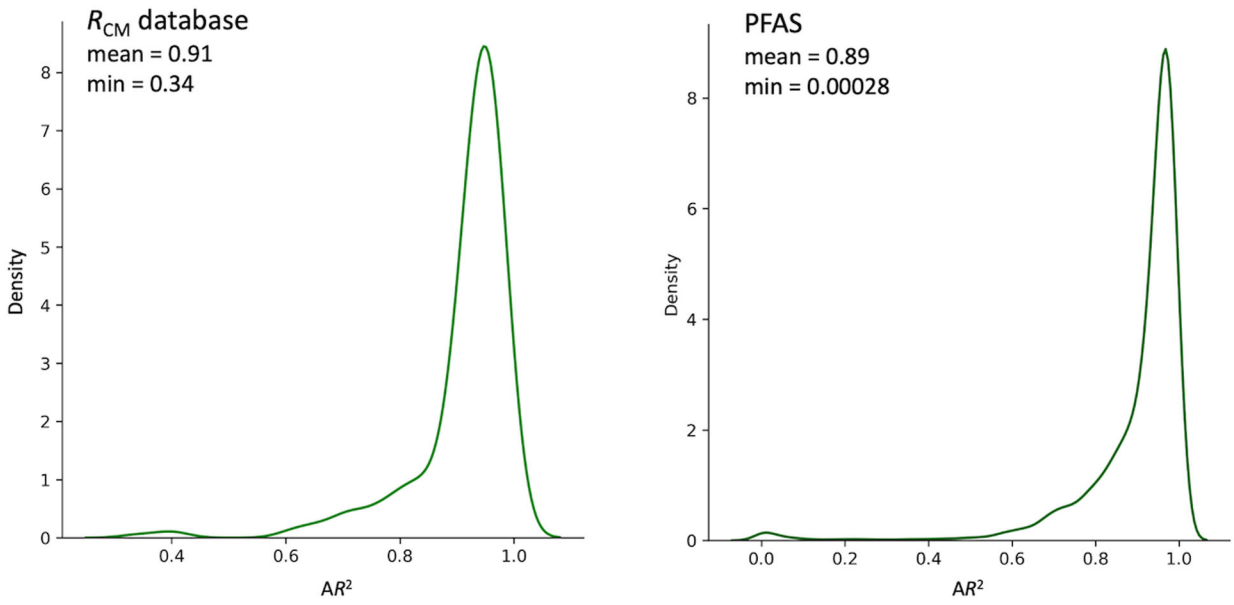

Figure 7:

Applicability R^2 (AR^2) calculated for the chemicals in the R_{CM} database and in PFASMASTER shown as a kernel density estimation plot. The applicability R^2 calculation is described in detail in the methods section of the manuscript. Briefly, the metric shows the similarity of a target compound to the "average" compound in the R_{CM} database in terms of physicochemical properties calculated using Mordred.

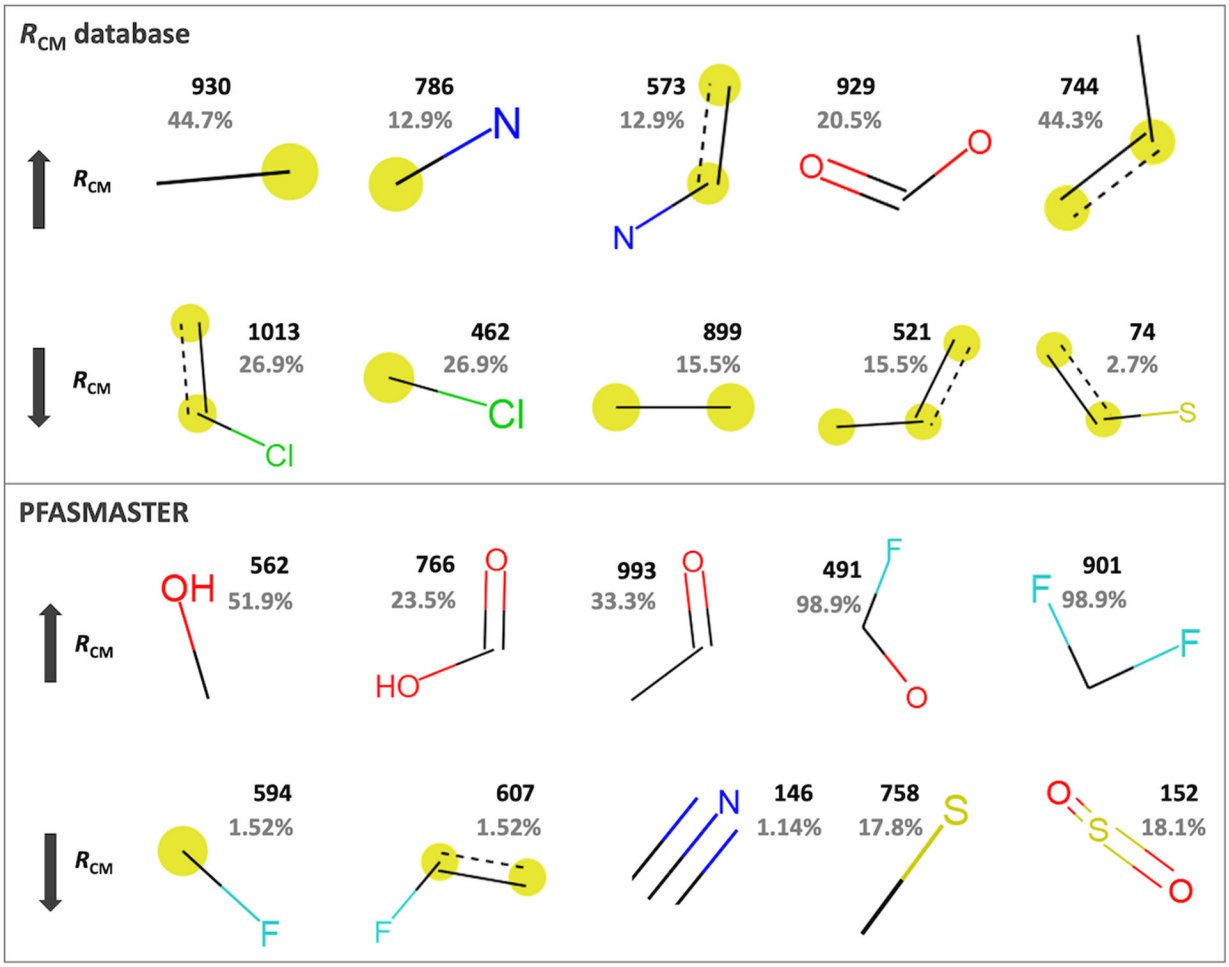

Figure 8:

Examples of RDKit bits (n steps = 2) that showed a significant ($p<0.05$) positive or negative association with R_{CM} for the chemicals in the R_{CM} database (n=260), and for the compounds from the PFAS database (n=7,982) based on the predicted R_{CM} values using the trained model. The bit number is shown in black letters and the frequency of that bit in chemicals in the database is shown in grey. Yellow circles represent aromatic carbons and dashed lines show the approximate electron distribution in aromatic bonds.

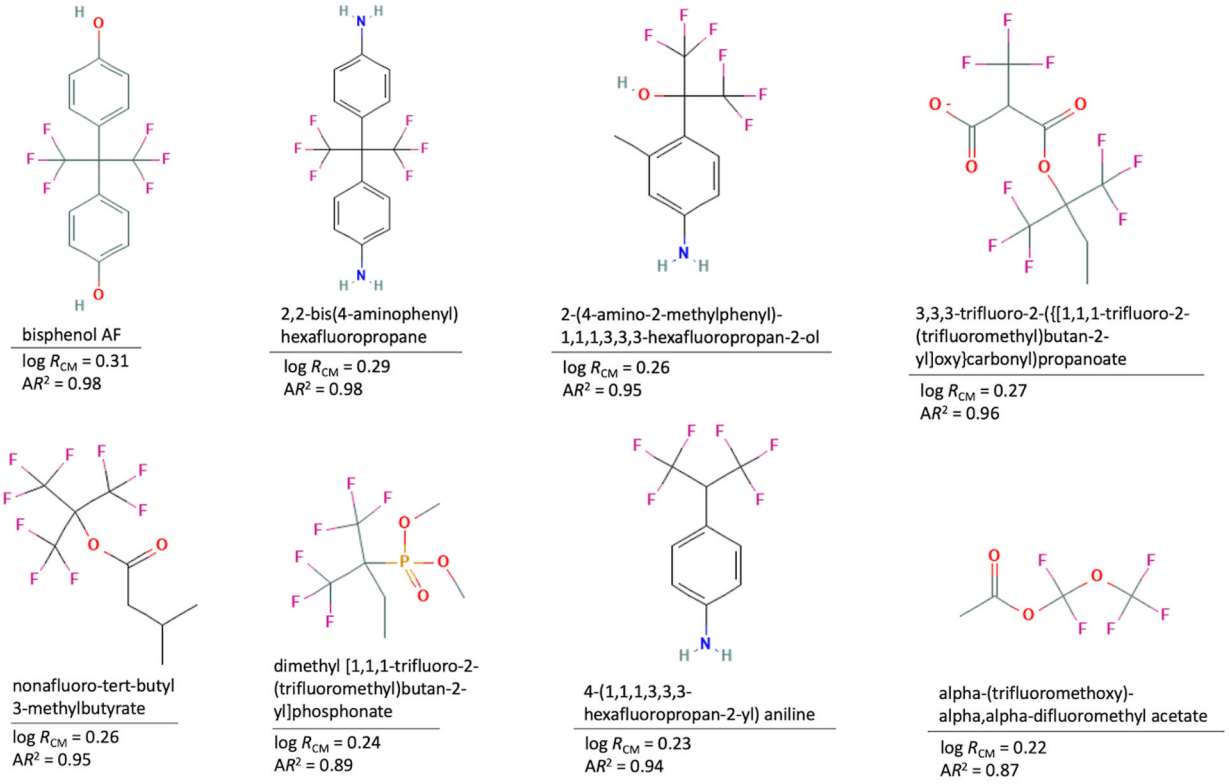

Figure 9:

Examples of PFAS that are expected to cross the placenta and preferentially partition to cord blood ($R_{CM} > 0$). The highest log R_{CM} observed was 0.31 for bisphenol AF indicating that cord blood concentrations of bisphenol AF are expected to be 3.1 times higher than the concentrations in maternal blood. All predicted R_{CM} values are presented in the Supplemental Spreadsheet 3.