

# Predicting Partition Coefficients of Short-Chain Chlorinated Paraffin Congeners by Combining COSMO-RS and Fragment Contribution Model Approaches

Satoshi Endo, Jort Hammer

Submitted date: 05/08/2020 • Posted date: 06/08/2020

Licence: CC BY-NC-ND 4.0

Citation information: Endo, Satoshi; Hammer, Jort (2020): Predicting Partition Coefficients of Short-Chain Chlorinated Paraffin Congeners by Combining COSMO-RS and Fragment Contribution Model Approaches. ChemRxiv. Preprint. <https://doi.org/10.26434/chemrxiv.12525656.v2>

Chlorinated paraffins (CPs) are highly complex mixtures of polychlorinated n-alkanes with differing chain lengths and chlorination patterns. Knowledge on physicochemical properties of individual congeners is limited but needed to understand their environmental fate and potential risks. This work combines a sophisticated but time-demanding quantum chemically based method COSMO-RS and a fast-running fragment contribution approach to establish models to predict partition coefficients of a large number of short-chain chlorinated paraffin (SCCP) congeners. Molecular fragments of a length of up to  $C_4$  in CP molecules were counted and used as explanatory variables to develop linear regression models for predicting COSMO-RS-calculated values. The resulting models can quickly provide COSMO-RS predictions for octanol–water ( $K_{ow}$ ), air–water ( $K_{aw}$ ), and octanol–air ( $K_{oa}$ ) partition coefficients of SCCP congeners with an accuracy of 0.1–0.3 log units root mean squared errors (RMSE). The model predictions for  $K_{ow}$  agree with experimental values for individual constitutional isomers within 1 log unit. The ranges of partition coefficients for each SCCP congener group were computed, which successfully reproduced experimental log  $K_{ow}$  ranges of industrial CP mixtures. As an application of the developed approach, the predicted  $K_{aw}$  and  $K_{oa}$  were plotted to evaluate the bioaccumulation potential of each SCCP congener group.

## File list (3)

Endo\_SCCPpartitioning\_20200622.pdf (1.22 MiB)

[view on ChemRxiv](#) • [download file](#)

Endo\_SCCPpartitioning\_SI\_20200622.pdf (2.67 MiB)

[view on ChemRxiv](#) • [download file](#)

Endo\_SCCPpartitioning\_SITables\_20200622.xlsx (87.61 KiB)

[view on ChemRxiv](#) • [download file](#)

# Predicting Partition Coefficients of Short-Chain Chlorinated Paraffin Congeners by Combining COSMO-RS and Fragment Contribution Model Approaches

*Satoshi Endo,\* Jort Hammer*

National Institute for Environmental Studies (NIES), Center for Health and Environmental Risk Research, Onogawa 16-2, 305-8506 Tsukuba, Ibaraki, Japan

\*Corresponding author

Satoshi Endo, Phone/Fax: ++81-29-850-2695, [endo.satoshi@nies.go.jp](mailto:endo.satoshi@nies.go.jp)

## Abstract

Chlorinated paraffins (CPs) are highly complex mixtures of polychlorinated *n*-alkanes with differing chain lengths and chlorination patterns. Knowledge on physicochemical properties of individual congeners is limited but needed to understand their environmental fate and potential risks. This work combines a sophisticated but time-demanding quantum chemically based method COSMO-RS and a fast-running fragment contribution approach to establish models to predict partition coefficients of a large number of short-chain chlorinated paraffin (SCCP) congeners. Molecular fragments of a length of up to C<sub>4</sub> in CP molecules were counted and used as explanatory variables to develop linear regression models for predicting COSMO-RS-calculated values. The resulting models can quickly provide COSMO-RS predictions for octanol–water ( $K_{ow}$ ), air–water ( $K_{aw}$ ), and octanol–air ( $K_{oa}$ ) partition coefficients of SCCP congeners with an accuracy of 0.1–0.3 log units root mean squared errors (RMSE). The model predictions for  $K_{ow}$  agree with experimental values for individual constitutional isomers within 1 log unit. The ranges of partition coefficients for each SCCP congener group were computed, which successfully reproduced experimental log  $K_{ow}$  ranges of industrial CP mixtures. As an application of the developed approach, the predicted  $K_{aw}$  and  $K_{oa}$  were plotted to evaluate the bioaccumulation potential of each SCCP congener group.

## Introduction

Chlorinated paraffins (CPs) are highly complex mixtures of polychlorinated *n*-alkanes with variable numbers of C and Cl atoms. Short-chain chlorinated paraffins (SCCPs, C<sub>10</sub>–C<sub>13</sub>) are considered persistent, bioaccumulative, and toxic and thus have been regulated under the Stockholm Convention since 2017.<sup>1</sup> Medium-chain CPs (MCCPs, C<sub>14</sub>–C<sub>17</sub>) and long-chain CPs (LCCPs, C<sub>18</sub>+) are not under regulation at the present time, although concerns have been raised, particularly for MCCPs, whether regulation should be implemented for these longer CPs.<sup>2</sup> CP products contain a considerable number of congeners with different molecular structures. To date, analytical methods are not available that fully resolve individual congeners from mixtures.<sup>3</sup> Environmental assessments for bulk CP mixtures often use average properties. However, once diluted in the environment, each congener behaves individually following its own properties. As has been learned from decades of studies on other halogenated organic pollutants such as polychlorinated dibenzo-*p*-dioxins and polychlorinated biphenyls, environmental behavior and toxicity are often highly congener-specific. Indeed, broad bands of CP signals from chromatographic analysis<sup>4</sup> suggest that the properties of congeners differ substantially.

To address the environmental fate and toxicity of individual CP congeners, their partitioning properties need to be understood. Experimental determination of such properties is only possible for a handful of congeners because the availability of pure analytical standards is currently limited. Computational methods may be the only possibility to provide congener-specific information. Among such prediction models available, empirical fit models may not be useful, as congener-specific experimental data are not sufficiently available to calibrate such models.

This study applies the quantum chemically based COSMO-RS theory<sup>5</sup> to predict partition coefficients of CP congeners. COSMO-RS can predict partition coefficients from the molecular structure alone without any additional empirical parameter. This approach could address partition coefficients of CP congeners with differing structures even including stereoisomers. Previous studies show that COSMO-RS can predict partition coefficients for chemicals of diverse structures (but no CPs) to the accuracy of < 1 log unit root-mean squared errors (RMSE) as compared to experimental data, including chemicals with multifunctional structure.<sup>6,7</sup> Relative values across chemicals are expected to be even more accurate because systematic errors are canceled.<sup>8</sup>

The problem of using COSMO-RS for predicting a large number of chemicals is the computational time needed for the quantum chemical calculation and the conformer selection. For example, it takes several hours to generate COSMO files, necessary to calculate partition coefficients, just for a single (stereo)isomer of C<sub>10</sub>Cl<sub>10</sub> using the supercomputer at the National Institute for Environmental Studies (HPE Apollo 2000, Intel Xeon Gold 6148 CPU, 40 CPU cores per each job). The computational time generally increases with the size of the molecule. Indeed, Glüge et al.<sup>9</sup> previously applied COSMO-RS to predict partition coefficients of CPs but provided predictions for only 4 structures per congener group, which are too few to address the variability of partition coefficients across congeners.

To enable the prediction of partition coefficients for hundreds of thousands of CP congeners, this study combines COSMO-RS with a fragment contribution model (FCM). An FCM counts the substructures (fragments) within the molecule and uses the fragment counts as descriptors for regression analysis. Such models have been widely adapted in the predictive model development of environmental properties.<sup>10-13</sup> FCMs are a linear model that can provide predictions with high speed and low electric energy consumption. In this work, we regress the COSMO-RS-predicted partition coefficients against CP's fragment counts to develop a model for predicting COSMO-RS predictions. Developing a model to predict the values that are output of another model might seem unmeaningful, because such a secondary model can only give less accurate predictions than the original model. However, such an approach is increasingly used in quantum chemistry applications where computational time is a hampering issue.<sup>14</sup> A secondary but fast-running fragment model could be useful particularly for CPs, and possibly other complex mixture components, because 1) experimental data for individual congeners are not available, 2) computation of the original model is too slow to cover the enormous number of congeners, and 3) the chemicals of concern are made up of relatively simple fragments and thus simple FCMs are expected to reproduce the predictions from a more sophisticated model well.

## Methods

**Method overview.** COSMO-RS-based FCMs for the log of octanol–water ( $K_{ow}$ ), air–water ( $K_{aw}$ ), and octanol–air ( $K_{oa}$ ) partition coefficients of SCCPs were developed by the following procedure. (1) The respective partition coefficients for a number of CP structures were calculated using the COSMO-RS method to generate training and validation sets. (2) FCMs with different combinations of fragments were calibrated using the training set. (3) Predictive performance of the calibrated FCMs was evaluated with the validation set. (4) Predictions by the FCMs were compared to available experimental data. (5) The FCMs were used to predict randomly generated SCCP congeners (1000 each for 52 congener groups) to demonstrate the variations of partition coefficients for SCCPs.

The CPs considered in this work are polychlorinated *n*-alkanes (i.e., no branching, no multiple bond). In this article, we refer to individual CP structures with different chain lengths and Cl-substitution patterns as “congeners”. A “congener group” collectively denotes the congeners with the same number of C and Cl atoms (i.e., isomers). Isomers of CPs include stereoisomers that have the same two-dimensional molecular structure but are not superimposable in the three-dimensional space because of the presence of chiral centers.

**COSMO-RS.** COSMO-RS calculates the chemical potential of solute in solution from quantum mechanics and statistical thermodynamics calculations and can thereby predict thermodynamic properties including partition coefficients.<sup>5</sup> For a given stereochemically specific congener, the molecular structure in the SDF format was entered into the COSMOconfX 4.3 software (COSMOlogic), which selected optimal conformers and generated their COSMO files using quantum chemistry program TURBOMOL 7.3 (COSMOlogic). These COSMO files were then used in COSMOthermX 19.0.4

(COSMOlogic, parameterization: BP\_TZVPD\_FINE\_19) to calculate  $K_{ow}$ ,  $K_{aw}$ , and  $K_{oa}$  at 25°C. Here, we calculated  $K_{ow}$  with wet octanol and  $K_{oa}$  with dry octanol. Note that the version of COSMOconfX used in this work sometimes returned structures that are stereochemically inconsistent with the original structure in the SDF (i.e., incorrect *R* or *S* configuration). This problem did not occur when we used the Windows version of COSMOconfX, switched off RDKit, and used only Balloon to generate initial candidate conformers.

The whole calculation procedure from COSMOconfX to COSMOthermX is consistent, although a slight difference in the calculated partition coefficient sometimes occurs when the initial input structure entered in COSMOconfX is in a different conformational state. We examined the extent of this “random error” using 10 starting conformations each for three arbitrarily chosen  $C_{10}$  congeners. The standard deviations for  $\log K_{ow}$ ,  $\log K_{aw}$ , and  $\log K_{oa}$  were on average 0.02, 0.14, and 0.12, respectively. These differences may represent the current precision of COSMOtherm predictions for CPs.

**Generation of training and validation sets.** In this work, we used “very” short to short-chain CPs ( $C_5$ – $C_{10}$ ) as training chemicals because computational time of the COSMOconfX optimization procedure increases with the size of molecule. By opting for short CPs, we were able to generate many congeners for model training. The validation set, in contrast, should comprise congeners that are relevant. We chose  $C_{10}$ – $C_{13}$ , thus SCCPs in this work. Calibrating and/or testing models for MCCPs and LCCPs would also be interesting but need much more time for calculations and was thus left for future work.

The training set consisted of 815 congeners—all 315 distinct isomers of  $C_5$  CPs and 100 randomly generated isomers for each of  $C_6$  to  $C_{10}$  CPs. In random generation, 0 to  $(2m + 2)$  H atoms of  $C_m$ -*n*-alkanes were randomly substituted with Cl atoms without any restriction. Here, all H atoms were considered distinct to also generate stereoisomers. Equivalent structures (i.e., superimposable by rotation) and enantiomers (i.e., mirror images) were removed because they show the identical partition coefficient value in reality, and COSMO-RS should give the same value in theory. Diastereomers, in contrast, can have different partitioning properties and thus are considered distinct congeners. The validation set consists of 120 CP congeners (30 for each of  $C_{10}$  to  $C_{13}$  CPs) that were also randomly generated. Codes were written in the *R* language<sup>15</sup> to create SMILES strings for all these congeners. SMILES was then converted to SDF format using OpenBabel,<sup>16</sup> which was then fed to COSMOconfX as described above.

**Fragment contribution models (FCMs).** Fragments in CP structures were counted using *R* with ChemmineR (3.38.0) and ChemmineOB (1.24.0) packages from Bioconductor.<sup>17</sup> Fragments with differing carbon-chain lengths, namely  $C_1$ ,  $C_2$ ,  $C_3$ , and  $C_4$  fragments were considered (Table S1 in the Supporting Information, SI). These fragments, respectively, have 7, 19, 64, and 220 types, out of which 0, 1, 10, and 67 types describe the diastereomeric patterns. All fragments and their SMARTS queries are given in Table S2. Using these fragments, four levels of models were generated. Level 1 model used only  $C_1$  fragments, and Level 2 model included  $C_1$  and  $C_2$  fragments. Levels 3 and 4 were

calibrated with  $C_1$  to  $C_3$  and  $C_1$  to  $C_4$ , respectively. Models were calibrated with least square multiple linear regression (MLR) with fragment counts as explanatory variables and COSMOtherm predictions as dependent variables. Since fragment counts are not fully independent and the contributions of some (or many) fragments to the partition coefficient can be insignificant, a forward and backward stepwise algorithm was used to select model variables (i.e., fragments). Akaike's Information Criterion (AIC) was considered the model evaluation metric.<sup>18</sup> Variable selection was first performed for the Level 1 model. Then, the selected  $C_1$  fragments were used as the initial variable set of the variable selection procedure for the Level 2 model, and so forth. To overcome a possible over-fitting problem, partial least squares regression (PLSR) was also performed using the selected Level 4 model fragments. The randomization test method was used to decide on the number of PLS components. All these statistical analyses were performed with *R* using functions such as `lm()`, `step()`, `plsr()`, and `selectNcomp()`.

**Predictions of partition coefficients for congener groups.** Using the FCMs calibrated with PLSR,  $\log K_{ow}$ ,  $\log K_{aw}$ , and  $\log K_{oa}$  for 1000 randomly generated isomers for each SCCP congener group ( $C_{10}$ – $C_{13}$ ,  $Cl_2$ – $Cl_{14}$ ) were predicted. Two methods were adapted to generate random isomers. In the first method, all H atoms were considered available for Cl substitution at the same likelihood. Second, all H atoms were available, but each C atom was able to carry a maximum of only one Cl atom. In other words, the first method allows double or triple Cl substitution, while the second does not. For the second case, congeners with the number of Cl > the number of C cannot be generated. Also, if Cl = C, then there is only one constitutional isomer (but with many stereoisomers). As for random generation of training and validation sets explained above, all substitution positions along the carbon-chain were considered distinct to account for stereoisomers. Duplications were allowed for random generation of 1000 isomers; this matters the most for  $C_{10}Cl_2$  group, which has only 30 constitutional isomers with 46 distinct structural isomers (i.e., 16 constitutional isomers have diastereomers). Duplication occurs increasingly rarely as the number of Cl approaches that of C. For example, 1000 random isomers of  $C_{10}Cl_{10}$  had only 10 duplications and 14 enantiomer pairs.

We are aware that existing studies have shown that Cl substitution patterns are not random in commercial CP mixtures. A recent study suggested that the first, second, and third carbons from an end of the chain and central carbons all have differing likelihood of chlorination.<sup>19</sup> Also, it has been known that chlorination occurs less likely to the neighbors of the carbon that is already chlorinated due to a steric effect,<sup>20,21</sup> which is also inferred by GC retention measurements for CP mixtures.<sup>4,22</sup> Nevertheless, in highly chlorinated CP mixtures, dichloro-substituted carbons and trichloromethyl groups have also been identified.<sup>19,23</sup> Since general rules for positions of Cl for CPs of different lengths and chlorination degree are still under investigation, we opted for the fully random and "one Cl per C" rules to generate congener sets for this work.

## Results and discussion

**FCM training and validation.** For all of  $\log K_{ow}$ ,  $\log K_{aw}$ , and  $\log K_{oa}$ , increasing the number of

fragments for MLR from Level 1 to Level 4 models improved the fitting quality, as indicated by  $R^2$ , root mean squared errors (RMSE), and AIC (Figures 1, S1, S2, Table S3). Hence, Level 4 model resulted in the best fit. It is interesting that  $C_4$  fragments do have statistically significant contributions to the partition coefficients, suggesting that the molecular interaction properties of CPs cannot fully be reduced to the shorter fragments, at least in the COSMO-RS calculation. In the variable selection procedure, about half (49-61%) of the total fragments were removed for Level 2 to 4 models. This is not surprising, because many fragments share common substructures and thus are interrelated. PLSR with the Level 4 fragments resulted in a similar fitting quality as compared to the least square MLR, although the PLSR has more restrictions (i.e., a lower degree of freedom) when deriving fitting coefficients. The good fit indicated by low RMSE (0.05, 0.12, and 0.09 for  $\log K_{ow}$ ,  $\log K_{aw}$ , and  $\log K_{oa}$ , respectively) with the Level 4 model or its PLSR version show that FCMs can accurately fit COSMO $therm$  calculated values for CPs. These RMSE values are similar to the precision of COSMO $therm$  for CPs mentioned in the Methods section and may thus be considered the best achievable fit. All resulting fitting coefficients are presented in Table S4. We note that some fragments that describe diastereometric structures were also significant.

External validation leads to the same conclusions as above. Thus, the Level 4 model showed the best statistics, and the statistics were better in order of  $\log K_{ow}$ ,  $\log K_{oa}$ , and  $\log K_{aw}$ . (Figures 1 and S3, Table S3). PLSR and the Level 4 model predicted the validation set equally well. While PLSR typically is considered more robust when the number of variables is large, it was just similar to the least square MLR-based Level 4 model. RMSE was 0.12, 0.29, and 0.21 for  $\log K_{ow}$ ,  $\log K_{aw}$ , and  $\log K_{oa}$ , respectively, being 2.2–2.4 times higher than RMSE for fitting. Use of the calibrated FCMs thus causes additional prediction errors of 0.1 to 0.3 RMSE in  $\log K$ 's of SCCPs, as compared to the direct use of COSMO $therm$ .

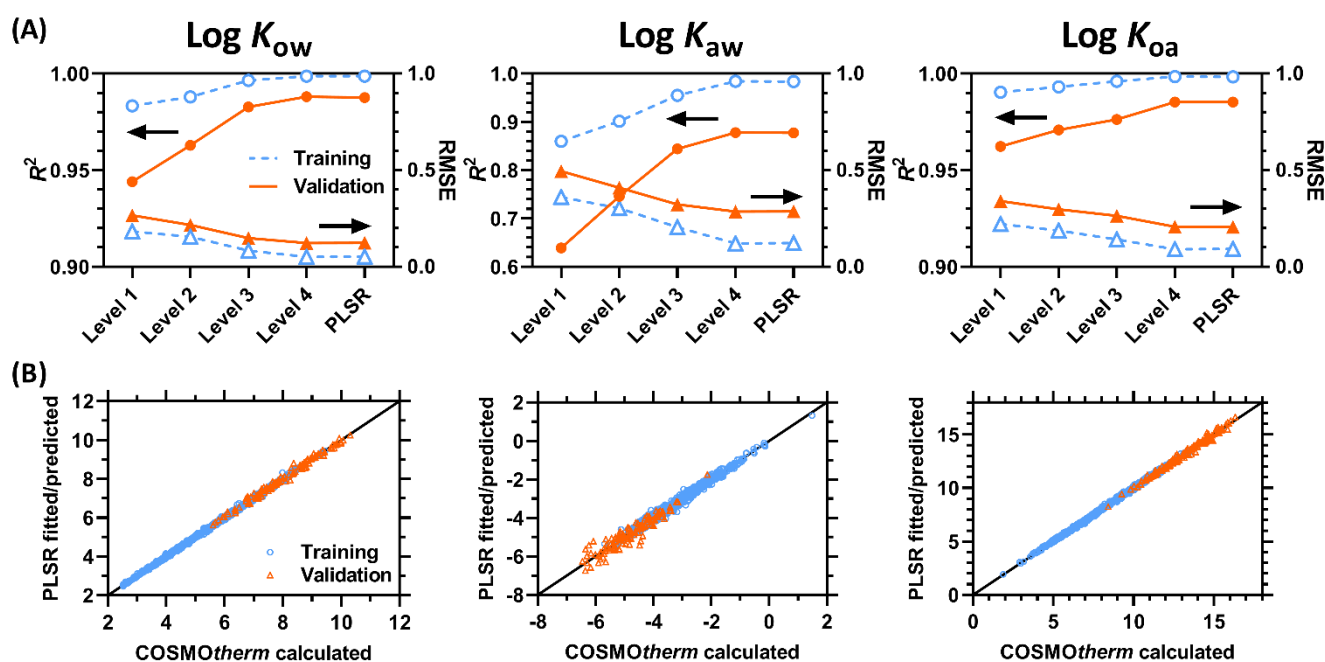


Figure 1. Statistics of model fitting and validation (A) and comparison of FCM (PLSR)-calculated values to training and validation data (B). Arrows indicate the axes that the data points refer to. Larger figures (Figures S1–S3) and a table with statistics for all models (Table S3) are presented in the SI.

**Fragment contributions to  $\log K$ 's.** The fact that the level 4 model performs the best suggests that the actual contribution of each C type (e.g.,  $-\text{CH}_2-$ ,  $-\text{CHCl}-$ ) to  $\log K$ 's depends on its neighboring structure. Nevertheless, lower level models may also be useful to illustrate the average contributions of the C types to  $\log K$ 's. For instance, the Level 1 model (with only  $\text{C}_1$  fragments) shows that the fragment contributions to  $\log K_{ow}$  and  $\log K_{oa}$  are fairly systematic (Figure S4): The  $-\text{CH}_2-$  increment increases  $\log K_{ow}$  and  $\log K_{oa}$ , while substituting H of  $-\text{CH}_2-$  or  $\text{CH}_3-$  with Cl also increases  $\log K_{ow}$  and  $\log K_{oa}$ . In contrast, the fragment contributions to  $\log K_{aw}$  are irregular. Substituting one H in  $-\text{CH}_2-$  with Cl to form  $-\text{CHCl}-$  decreases  $\log K_{aw}$ , but further substitution to  $-\text{CCl}_2-$  would not change  $\log K_{aw}$ . Similarly, Cl-substitution of  $\text{CH}_3-$  to  $\text{CH}_2\text{Cl}-$  decreases  $\log K_{aw}$ , but further substitution to  $\text{CHCl}_2-$  has no influence, and that to  $\text{CCl}_3-$  even increases  $\log K_{aw}$ .

**Comparison to experimental data.** There are some experimentally determined  $\log K_{ow}$  and  $\log K_{aw}$  for specific constitutional isomers in the literature. The predictions by the FCM (PLSR-calibrated) agree with the literature data for  $\log K_{ow}$  within 1 log unit difference (Figure 2). The FCM tends to overpredict  $\log K_{ow}$  of CP congeners with five or more chlorinated C atoms. The predictions by the original COSMOtherm deviate from the experimental data to a similar extent. Thus, the observed overpredictions for some  $\log K_{ow}$  data should be related to the inaccuracy in the original COSMOtherm calculations or due to experimental errors. The cited experimental data were derived from HPLC retention measurements using 10 hydrophobic aromatic compounds as calibration



compounds. While this is a standard approach, the resulting  $K_{ow}$  may not be as accurate for aliphatic chemicals with appreciable polarity like CPs as for hydrophobic aromatic compounds.

The two congener-specific experimental  $\log K_{aw}$  are underpredicted by the FCM by 1–1.5 log units. The original COSMOtherm predictions agree better with the experimental data in this case. As Figure 2 shows, the predictions for  $K_{aw}$  by FCM and COSMOtherm differ by ca 1 log unit, which is close to the maximal error (0.94 log units) found in the model validation presented above. The reason for the model disagreement specifically for 1,2,9,10- $C_{10}Cl_4$  and 1,2,10,11- $C_{10}Cl_4$  is unknown, but we speculate that an extended sequence of non-Cl-substituted  $-CH_2-$  units rarely occurs in our random isomer generation for the training set and thus may have been under-represented in the model training. Indeed, our training set contained only two congeners with a  $-CH_2-CH_2-CH_2-CH_2-$  fragment and none with a longer  $CH_2$  chain. That said, the experimental  $K_{aw}$  value for 1,2,9,10- $C_{10}Cl_4$  could also be somewhat too high, as  $\log K_{aw}$  of  $-2$  in combination with  $\log K_{ow}$  of  $5$  would result in  $\log K_{oa}$  of  $7$  (via  $K_{oa} = K_{ow}/K_{aw}$ ) and this is even smaller than an experimentally measured log hexadecane–air partition coefficient ( $L$ ) of  $8.4$  for 1,2,9,10- $C_{10}Cl_4$ .<sup>22</sup>  $K_{oa} \gtrsim L$  is generally expected because CPs interact with octanol via additional polar interactions that do not occur with apolar hexadecane.

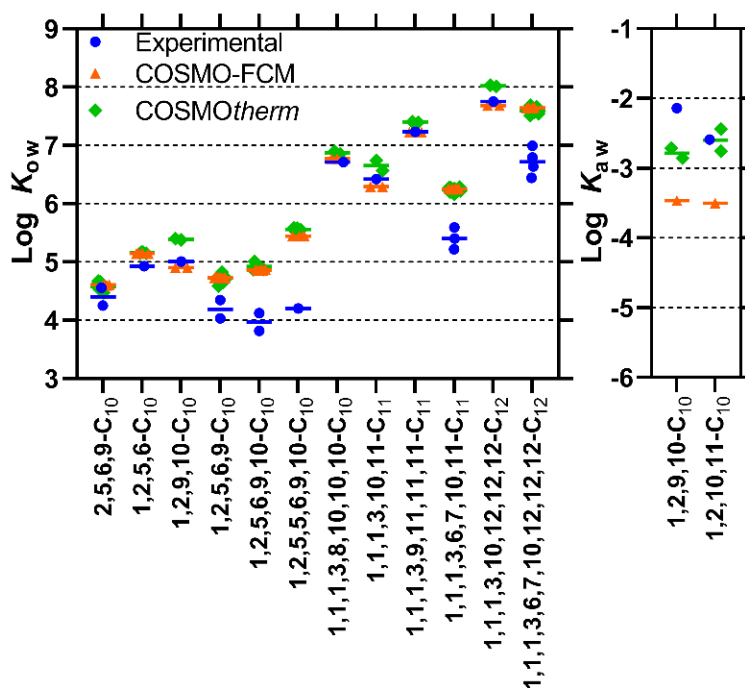


Figure 2. Comparison of predicted with experimental  $K$  values. Predictions were derived from COSMO-RS based FCMs (with PLSR calibration) as well as directly from COSMOtherm software.  $K_{ow}$  data are from Hilger et al.<sup>24</sup>  $K_{aw}$  data are from Drouillard et al.<sup>25</sup> for 23°C. There are multiple data points both for predictions and experimental data because of the presence of diastereomers. Bars indicate the mean.

253 **Distributions of  $K_{ow}$ ,  $K_{aw}$ , and  $K_{oa}$  for SCCP congener groups.** Using the FCM based on PLSR  
254 calibration, log  $K$ 's for 1000 isomers per congener group were predicted. These predictions were  
255 used to estimate the distributions of  $K_{ow}$ ,  $K_{aw}$ , and  $K_{oa}$  for SCCP congener groups (Figures 3, S5; here  
256 double/triple Cl is not allowed). The 2.5, 25, 50 (median), 75, and 97.5%iles of log  $K$ 's for each SCCP  
257 congener group are presented in Tables S5.

258 Log  $K_{ow}$  and log  $K_{oa}$  for each congener group are within a relatively narrow range (1 to 2 log  
259 units), whereas log  $K_{aw}$  for each congener group spreads over 1.5 to 3 log units. The median log  $K_{ow}$   
260 values of different SCCP congener groups range over 4 log units (5.0 to 8.9) and log  $K_{aw}$  also over 4  
261 log units (–5.7 to –1.6), whereas the median log  $K_{oa}$  spans over 8 log units (6.7 to 14.8).

262 It is interesting that the medians of log  $K_{ow}$ ,  $K_{aw}$ , and  $K_{oa}$  show different dependence on the  
263 numbers of C and Cl. All three log  $K$ 's are linearly dependent on the number of C, although the slopes  
264 differ depending on the partitioning phases and partially on the number of Cl (Figure S6). In contrast,  
265 dependence on the number of Cl is nonlinear (Figure 3; more clearly in Figure S7). Log  $K_{ow}$  is fairly  
266 constant from Cl<sub>2</sub> to ~Cl<sub>5</sub>, above which it increases with ca 0.35 log units/C. Log  $K_{aw}$  has the opposite  
267 trend; it decreases from Cl<sub>2</sub> to ~Cl<sub>10</sub> by 2.5–3.5 log units and thereafter stays nearly the same. Log  $K_{oa}$   
268 monotonically increases but in a concave downward shape. The increase is ca 0.8 log unit/Cl from Cl<sub>2</sub>  
269 to Cl<sub>3</sub> whereas only 0.4 log units/Cl from C<sub>13</sub> to C<sub>14</sub>.

270 We also derived the distributions of partition coefficients for CP congeners without double  
271 and triple Cl substitutions (Figure S8, Table S6). The distribution peaks are sometimes slightly sharper,  
272 but overall, the results are just similar to the distributions of CP congeners including double/triple Cl  
273 substitutions. The median for each congener group is different by only 0.13 log unit on average and  
274 by 0.39 at most. The similarity between the cases with and without double/triple Cl is expected for  
275 low-chlorinated congeners (Cl<sub>2</sub>–Cl<sub>4</sub>), because random generation forms limited numbers of double  
276 and triple substitutions, even if allowed. The similarity for higher chlorinated congeners is interesting  
277 and suggests that this difference is not important for partition coefficients when the congener groups  
278 are considered as a whole.

279

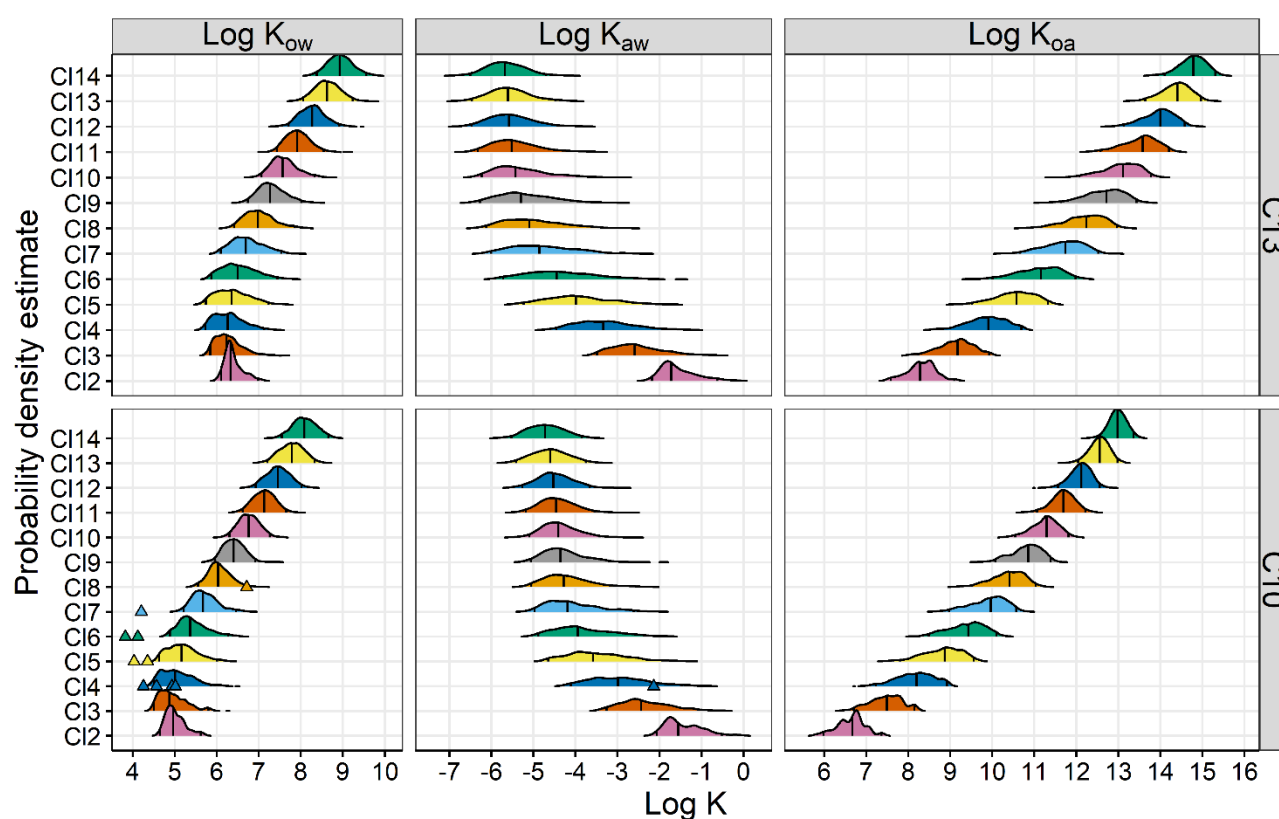


Figure 3. Kernel density estimates resulting from 1000 structures (with double and triple Cl substitution allowed) for each molecular formula (predicted by PLSR model). Vertical lines indicate the 2.5, 50 (median), and 97.5%iles. Data points are experimental data from Hilger et al.<sup>24</sup> and Drouillard et al.<sup>25</sup> for specific isomers. Plots with more congener groups and without double and triple Cl substitutions are shown in Figures S5 and S8.

**Predicting log  $K_{ow}$  of SCCP mixtures.** The log  $K_{ow}$  distributions predicted above for all relevant SCCP congener groups were used in combination with the compositions of SCCP mixtures experimentally derived from Yuan et al.<sup>26</sup> to predict log  $K_{ow}$  ranges of bulk CP mixtures (Figure 4; more plots in Figures S9, S10). Here, the predicted log  $K_{ow}$  distributions for congener groups were weighted by their relative abundance (i.e., mole fractions) in the mixture and were then summed. The results agreed with the experimental data from Renberg et al.,<sup>27</sup> who used retentions on thin layer chromatography to estimate the ranges of log  $K_{ow}$  for CP mixtures. The lower bound of the experimental data agrees with the predicted 2.5%ile and the upper bound with the predicted 97.5%ile within 0.84 log units (Table S7). These results serve as additional validation of COSMOtherm and the FCMs for predicting log  $K_{ow}$  of CPs.

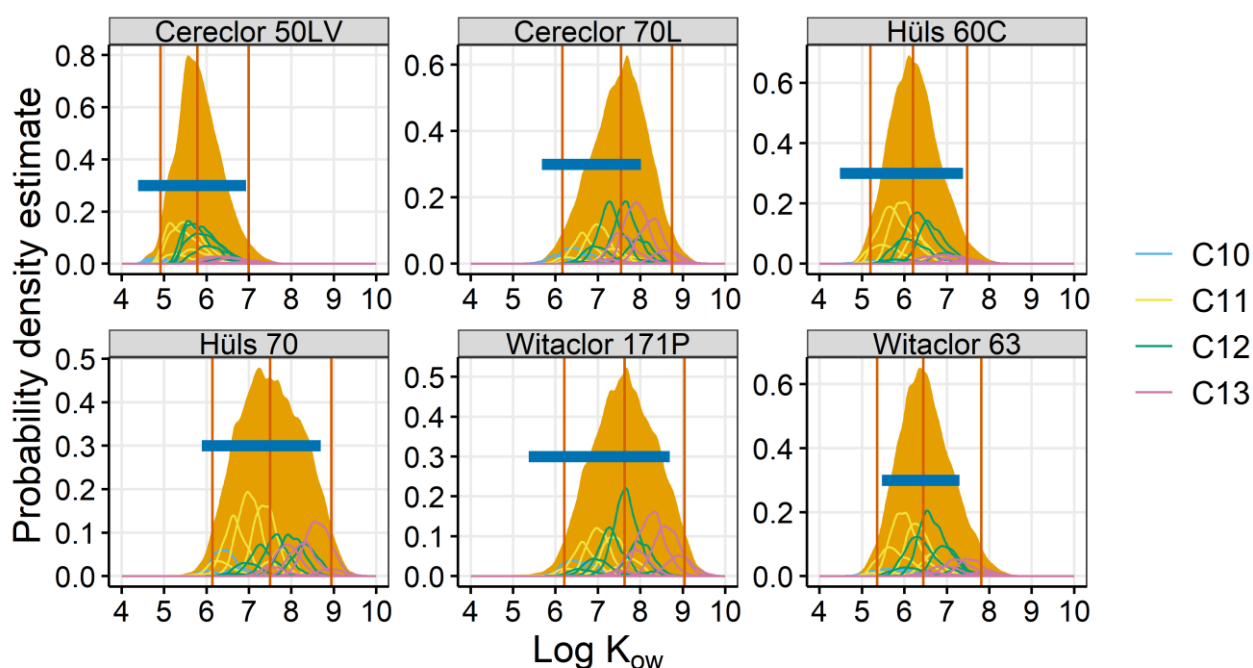


Figure 4. Comparison of predicted distributions of  $\log K_{ow}$  for CP mixtures (filled curves) and experimental data from Renberg et al.<sup>27</sup> (horizontal bars). The 2.5, 50, and 97.5%iles of the predications for mixtures (vertical lines) and the predictions for each congener groups (unfilled curves) are also shown. The predictions were derived from the FCMs (PLSR, double/triple CI allowed).

**Implications.** This study presents, for the first time, a time-efficient method to predict partition coefficients for a large number of CP congeners on the basis of quantum chemical calculations. We provided the ranges of partition coefficients for CP congener groups and bulk CP mixtures, grounded on the predictions for individual congeners. These new pieces of information should improve our understanding on the environmental fate of CPs. As an example, SCCP congeners were plotted in the chemical space that indicates the Arctic bioaccumulation potential using predicted  $\log K_{aw}$  and  $\log K_{oa}$ , following the approach by Czub et al.<sup>28</sup> and Brown and Wania<sup>29</sup> (Figure 5; plots for each SCCP congener group is shown in Figure S10). Figure 5 shows that relatively low chlorinated ( $Cl_2$ – $Cl_6$ ) SCCPs fall into the chemical space where high Arctic bioaccumulation is expected, assuming perfect persistence. In contrast, SCCPs with relatively high molecular weight ( $C + Cl \geq 20$ ) do not fall in this zone. Previously, Gawor and Wania<sup>30</sup> presented various chemical space plots for CPs using  $\log K_{aw}$  and  $\log K_{oa}$  predicted by ACD/ADME Suite prediction tools and came up with conclusions that are in part similar to those in this work. It would be interesting to repeat their analysis with the predicted partition coefficients from this work, which is however beyond the scope of this article.

The presented approach of course has room for further improvement. First, the current FCMs have been calibrated with relatively short CPs ( $\leq C_{10}$ ) and thus would have to be extrapolated for M/LCCPs. Extension to M/LCCPs requires lengthy COSMOconfX calculations for long molecules, which

will be conducted as a next step. Second, while the current study demonstrated good model predictions for log  $K_{ow}$  of a dozen of individual constitutional isomers and six bulk SCCP mixtures, validation for more of specific (stereo)isomers and other partition phases would be desirable. This statement applies both to the original COSMO-RS approach and the FCMs presented here. Availability of isomer-specific CP standards is being improved and more data are expected in the future (e.g., ref 22). Third, the current work used randomly generated congeners from all Cl substitution patterns or excluding double and triple Cl substitutions to represent the congener composition of each CP congener group, but this is a first approximation. As more and more knowledge regarding Cl substitution patterns in the bulk CP mixtures is becoming available,<sup>19,23,31</sup> congener compositions used for the prediction of partition coefficients could be elaborated further.

This study demonstrated that the approach that combines COSMO-RS and FCM methods can provide accurate predictions for SCCP partitioning coefficients. As the most time-consuming COSMO*confX* step that generates COSMO files has been completed for a number of CP congeners, it is possible to run COSMO*therm* and derive new FCMs quickly for other partition coefficients or other properties of CPs that are related to the chemical potential in solvent. Our approach may be useful for other highly complex mixtures as well, as partitioning properties of complex mixtures are generally difficult to determine both experimentally and computationally.

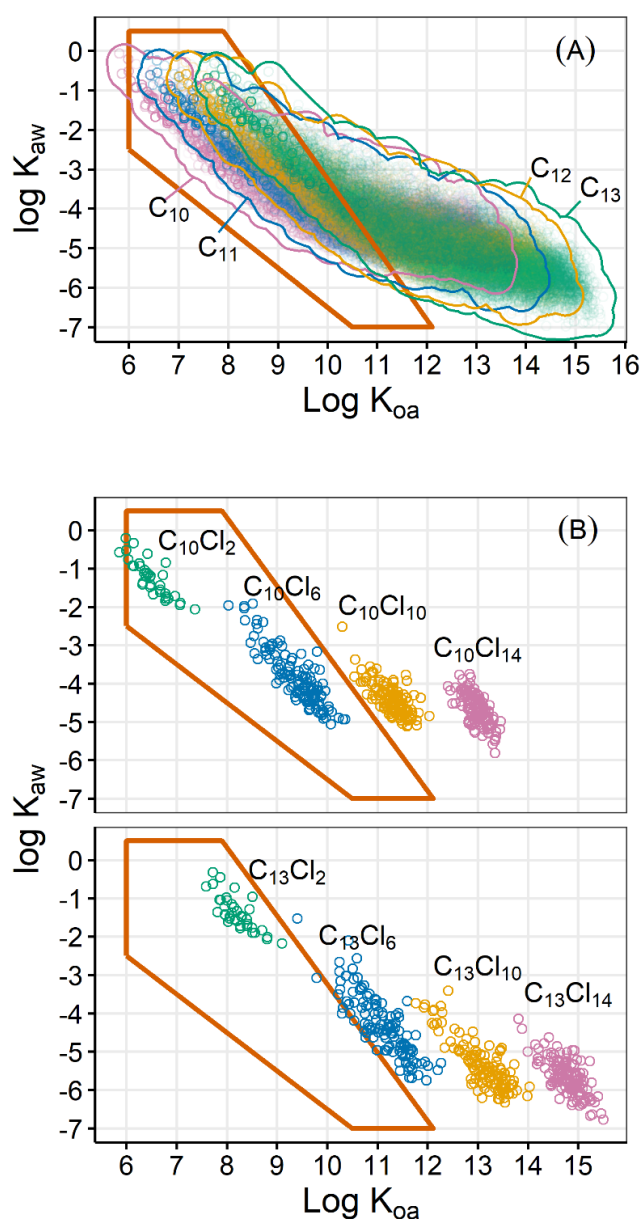


Figure 5. Chemical space plots for (A) all and (B) selected SCCP congener groups. The chemical space for a high Arctic contamination and bioaccumulation potential AC-BAP (>10%, 70 days)<sup>28</sup> was enclosed with lines, as in ref 29.

## Associated Content

### Supporting Information

Additional figures and tables for the results of model fitting and validation, model predictions for  $\log K'$ s, distributions of  $\log K_{ow}$ , and chemical space plots. This material is available free of charge via the Internet at <http://pubs.acs.org>.

## Author Information

Corresponding author

Satoshi Endo

Phone/Fax: ++81-29-850-2695

endo.satoshi@nies.go.jp

ORCID: 0000-0001-8702-1602

## Notes

The authors declare no competing financial interest.

## Author contributions

Study design: SE. COSMO-RS calculations: SE, JH. Statistical analysis: SE. Data evaluation: SE. Drafting of manuscript: SE. Revising of manuscript: SE, JH.

## Acknowledgements

We thank Frank Wania, Trevor Brown, and Kai-Uwe Goss for their valuable comments on the manuscript. COSMOconfX and TURBOMOL calculations were performed with the NIES supercomputer system. This work was supported by the Environment Research and Technology Development Fund (SII-3-1) of the Environmental Restoration and Conservation Agency, Japan.

## References

1. UNEP, Decision SC-8/11: Listing of short-chain chlorinated paraffins. 2017, UNEP/POPS/COP.8/SC-8/11.
2. Glüge, J.; Schinkel, L.; Hungerbühler, K.; Cariou, R.; Bogdal, C., Environmental risks of medium-chain chlorinated paraffins (MCCPs): A Review. *Environ. Sci. Technol.* **2018**, *52*, (12), 6743-6760.
3. van Mourik, L. M.; Lava, R.; O'Brien, J.; Leonards, P. E. G.; de Boer, J.; Ricci, M., The underlying challenges that arise when analysing short-chain chlorinated paraffins in environmental matrices. *J. Chromatogr. A* **2020**, *1610*, 460550.
4. Korytar, P.; Parera, J.; Leonards, P. E.; Santos, F. J.; de Boer, J.; Brinkman, U. A., Characterization of polychlorinated *n*-alkanes using comprehensive two-dimensional gas chromatography--electron-capture negative ionisation time-of-flight mass spectrometry. *J. Chromatogr. A* **2005**, *1086*, (1-2), 71-82.
5. Klamt, A., Conductor-like screening model for real solvents: A new approach to the quantitative calculation of solvation phenomena. *J. Phys. Chem.* **1995**, *99*, (7), 2224-2235.
6. Stenzel, A.; Goss, K.-U.; Endo, S., Prediction of partition coefficients for complex environmental contaminants: Validation of COSMOtherm, ABSOLV, and SPARC. *Environ. Toxicol. Chem.* **2014**, *33*, (7), 1537-1543.

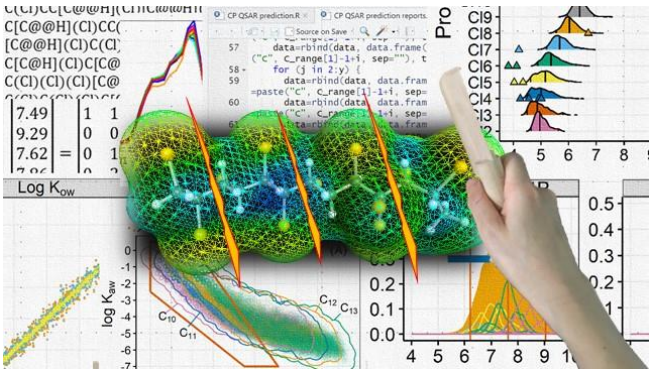
- 390 7. Loschen, C.; Reinisch, J.; Klamt, A., COSMO-RS based predictions for the SAMPL6 logP  
391 challenge. *J. Comput. Aided Mol. Des.* **2020**, *34*, (4), 385-392.
- 392 8. Goss, K.-U.; Arp, H. P. H.; Bronner, G.; Niederer, C., Partition behavior of  
393 hexachlorocyclohexane isomers. *J. Chem. Eng. Data* **2008**, *53*, (3), 750-754.
- 394 9. Glüge, J.; Bogdal, C.; Scheringer, M.; Buser, A. M.; Hungerbühler, K., Calculation of  
395 physicochemical properties for short- and medium-chain chlorinated paraffins. *J. Phys.*  
396 *Chem. Ref. Data* **2013**, *42*, (2), 023103.
- 397 10. Meylan, W. M.; Howard, P. H., Atom/fragment contribution method for estimating octanol-  
398 water partition coefficients. *J. Pharm. Sci.* **1995**, *84*, (1), 83-92.
- 399 11. Brown, T. N.; Arnot, J. A.; Wania, F., Iterative fragment selection: a group contribution  
400 approach to predicting fish biotransformation half-lives. *Environ. Sci. Technol.* **2012**, *46*, (15),  
401 8253-8260.
- 402 12. OECD, Guidance document on the validation of (quantitative) structure-activity relationships  
403 [(Q)SAR] models. 2007, ENV/JM/MONO(2007)2.
- 404 13. Dearden, J. C.; Cronin, M. T.; Kaiser, K. L., How not to develop a quantitative structure-activity  
405 or structure-property relationship (QSAR/QSPR). *SAR QSAR Environ. Res.* **2009**, *20*, (3-4),  
406 241-266.
- 407 14. Ramakrishnan, R.; von Lilienfeld, O. A., Machine learning, quantum chemistry, and chemical  
408 space. In *Reviews in Computational Chemistry, Vol 30*, Parrill, A. L.; Lipkowitz, K. B., Eds.  
409 Wiley-Blackwell: Malden, 2017; Vol. 30, pp 225-256.
- 410 15. R Core Team, *R: A language and environment for statistical computing*. R Foundation for  
411 Statistical Computing, Vienna, Austria. <https://www.R-project.org/>. 2019.
- 412 16. O'Boyle, N. M.; Banck, M.; James, C. A.; Morley, C.; Vandermeersch, T.; Hutchison, G. R.,  
413 Open Babel: An open chemical toolbox. *J. Cheminformatics* **2011**, *3*, (1), 33.
- 414 17. Cao, Y.; Charisi, A.; Cheng, L. C.; Jiang, T.; Girke, T., ChemmineR: a compound mining  
415 framework for R. *Bioinformatics* **2008**, *24*, (15), 1733-1734.
- 416 18. Burnham, K. P.; Anderson, D. R., Multimodel Inference. *Sociol. Methods Res.* **2016**, *33*, (2),  
417 261-304.
- 418 19. Yuan, B.; Lysak, D. H.; Soong, R.; Haddad, A.; Hisatsune, A.; Moser, A.; Golotvin, S.;  
419 Argyropoulos, D.; Simpson, A. J.; Muir, D. C. G., Chlorines are not evenly substituted in  
420 chlorinated paraffins: A predicted nmr pattern matching framework for isomeric  
421 discrimination in complex contaminant mixtures. *Environ. Sci. Technol. Letters* **2020**.  
422 doi.org/10.1021/acs.estlett.0c00244.
- 423 20. Fredricks, P. S.; Tedder, J. M., 28. Free-radical substitution in aliphatic compounds. Part II.  
424 Halogenation of the *n*-butyl halides. *J. Chem. Soc. (Resumed)* **1960**, 144-150.
- 425 21. Colebourne, N.; Stern, E. S., 660. The chlorination of some *n*-alkanes and alkyl chlorides. *J.*  
426 *Chem. Soc. (Resumed)* **1965**, 3599-3605.
- 427 22. Hammer, J.; Matsukami, H.; Endo, S., Congener-specific partition properties of chlorinated



- paraffins evaluated with COSMOtherm and GC-retention indices. (Manuscript in preparation).
23. Sprengel, J.; Wiedmaier-Czerny, N.; Vetter, W., Characterization of single chain length chlorinated paraffin mixtures with nuclear magnetic resonance spectroscopy (NMR). *Chemosphere* **2019**, *228*, 762-768.
  24. Hilger, B.; Fromme, H.; Volkel, W.; Coelhan, M., Effects of chain length, chlorination degree, and structure on the octanol-water partition coefficients of polychlorinated *n*-alkanes. *Environ Sci Technol* **2011**, *45*, (7), 2842-9.
  25. Drouillard, K. G.; Tomy, G. T.; Muir, D. C. G.; Friesen, K. J., Volatility of chlorinated *n*-alkanes (C-10-C-12): Vapor pressures and Henry's law constants. *Environ. Toxicol. Chem.* **1998**, *17*, (7), 1252-1260.
  26. Yuan, B.; Bogdal, C.; Berger, U.; MacLeod, M.; Gebbink, W. A.; Alsberg, T.; de Wit, C. A., Quantifying short-chain chlorinated paraffin congener groups. *Environ. Sci. Technol.* **2017**, *51*, (18), 10633-10641.
  27. Renberg, L.; Sundstrom, G.; Sundhnygard, K., Partition coefficients of organic chemicals derived from reversed phase thin-layer chromatography: Evaluation of methods and application on phosphate esters, polychlorinated paraffins and some PCB-substitutes. *Chemosphere* **1980**, *9*, (11), 683-691.
  28. Czub, G.; Wania, F.; McLachlan, M. S., Combining long-range transport and bioaccumulation considerations to identify potential Arctic contaminants. *Environ. Sci. Technol.* **2008**, *42*, (10), 3704-3709.
  29. Brown, T. N.; Wania, F., Screening chemicals for the potential to be persistent organic pollutants: A case study of Arctic contaminants. *Environ. Sci. Technol.* **2008**, *42*, (14), 5202-5209.
  30. Gawor, A.; Wania, F., Using quantitative structural property relationships, chemical fate models, and the chemical partitioning space to investigate the potential for long range transport and bioaccumulation of complex halogenated chemical mixtures. *Environ. Sci. Process Impacts* **2013**, *15*, (9), 1671-1684.
  31. Jensen, S. R.; Brown, W. A.; Heath, E.; Cooper, D. G., Characterization of polychlorinated alkane mixtures-A Monte Carlo modeling approach. *Biodegradation* **2007**, *18*, (6), 703-717.

460 TOC

461  
462



Endo\_SCCPpartitioning\_20200622.pdf (1.22 MiB)

[view on ChemRxiv](#) • [download file](#)

---

# Supporting Information for

## “Predicting partition coefficients of short-chain chlorinated paraffin congeners by combining COSMO-RS and fragment contribution model approaches”

Satoshi Endo,\* Jort Hammer

National Institute for Environmental Studies (NIES), Center for Health and Environmental Risk Research, Onogawa 16-2, 305-8506 Tsukuba, Ibaraki, Japan

\*Corresponding author, Satoshi Endo, Phone/Fax: ++81-29-850-2695, [endo.satoshi@nies.go.jp](mailto:endo.satoshi@nies.go.jp)

### Contents

Figure S1. Fragment model fitting to the training set.

Figure S2. Residual plots for fragment model fitting to COSMOtherm calculated values.

Figure S3. Fragment model predictions for the validation set

Figure S4. Fragment contributions to log  $K'$ s in the Level 1 model.

Figure S5. Kernel density estimates resulting from 1000 isomer structures for each congener group (predicted by PLSR model, with double and triple Cl substitutions allowed).

Figure S6. Dependence of the median log  $K'$ s for SCCP congener groups on the numbers of C.

Figure S7. Dependence of the median log  $K'$ s for SCCP congener groups on the numbers of Cl.

Figure S8. Kernel density estimates resulting from 1000 isomer structures for each congener group (predicted by PLSR model).

Figure S9. Distributions of log  $K_{ow}$  for CP mixtures (I).

Figure S10. Distributions of log  $K_{ow}$  for CP mixtures (II).

Figure S11. Chemical space plot.

Table S1. Types of fragments used in the model development.

Table S2. Fragments considered and their SMARTS queries. (In the Excel file)

Table S3. Results of model fitting and validation.

Table S4. Resulting fitting coefficients of FCMs. (In the Excel file)

Table S5. Quantiles of log  $K'$ s for each congener group (predicted by PLSR model, with double/triple Cl substitutions allowed). (in the Excel file)

Table S6. Quantiles of log  $K'$ s for each congener group (predicted by PLSR model, with double/triple Cl substitutions not allowed). (In the Excel file)

Table S7. Predicted and experimental log  $K_{ow}$  ranges for CP mixtures. Experimental data are from Renberg et al.

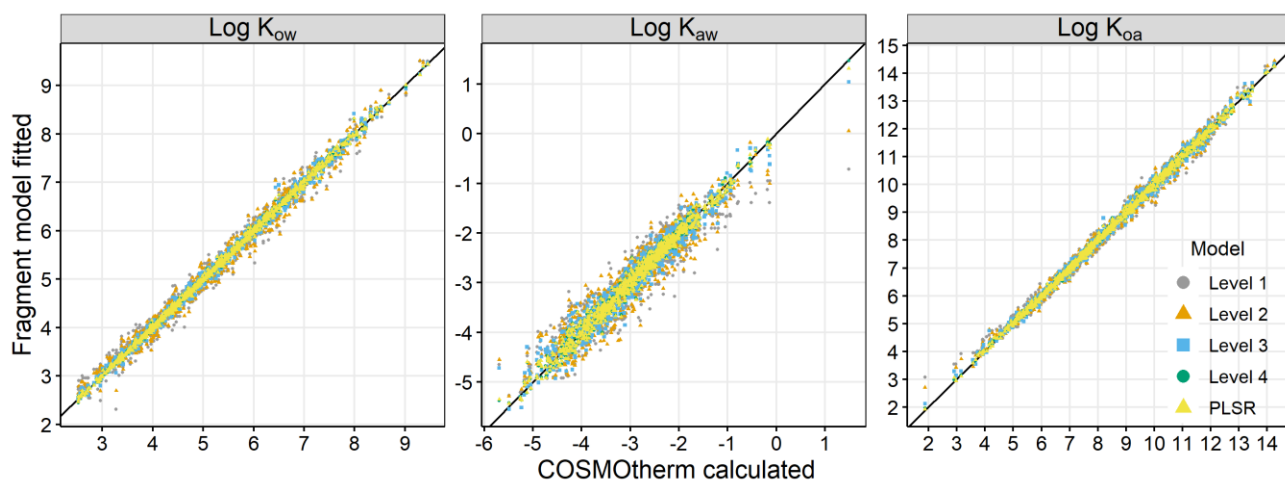


Figure S1. Fragment model fitting to the training set.

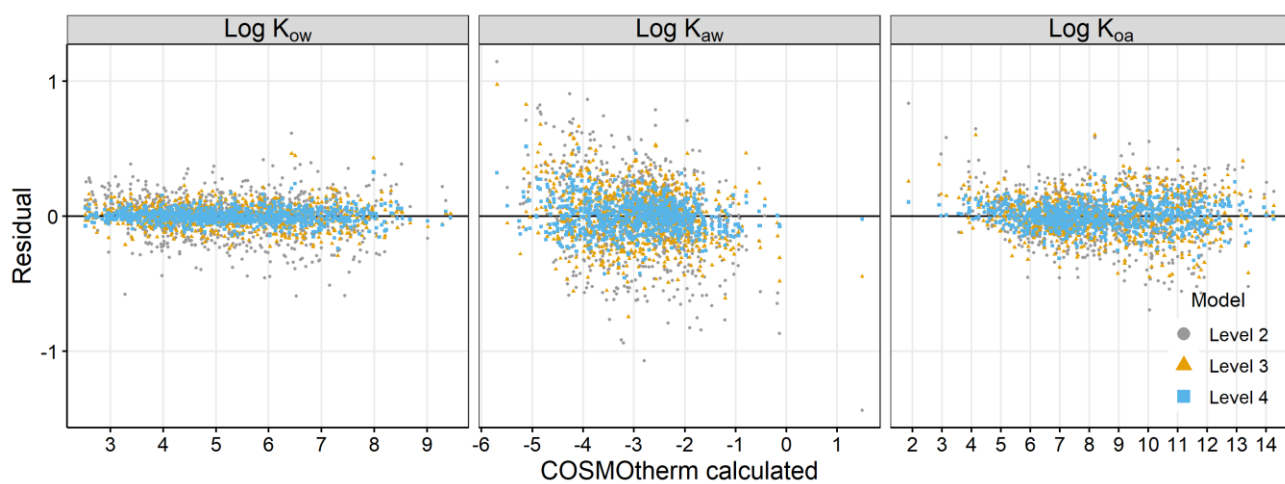


Figure S2. Residual plots for fragment model fitting to COSMOtherm calculated values.

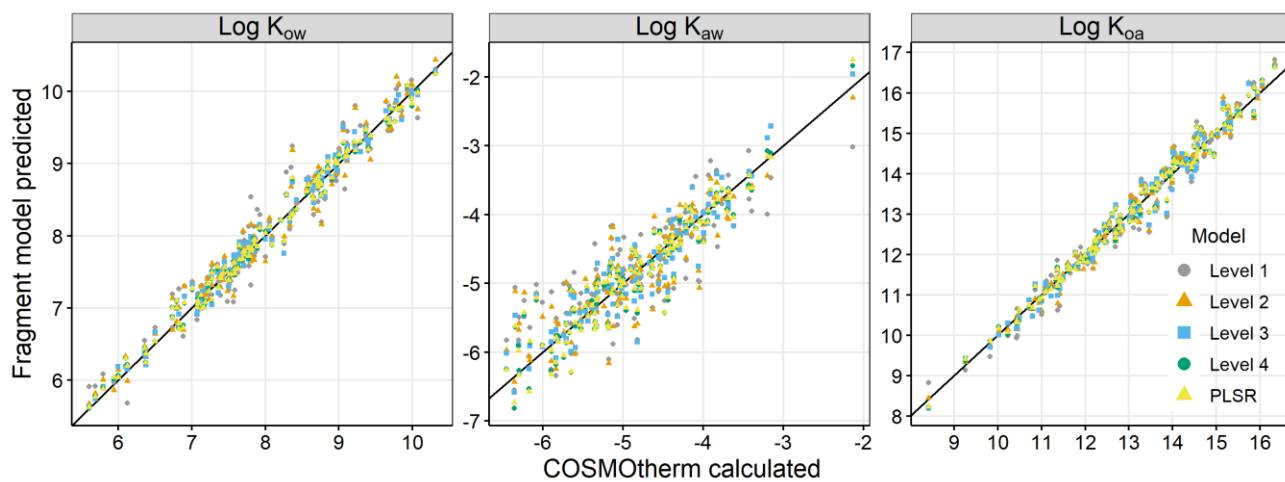


Figure S3. Fragment model predictions for the validation set

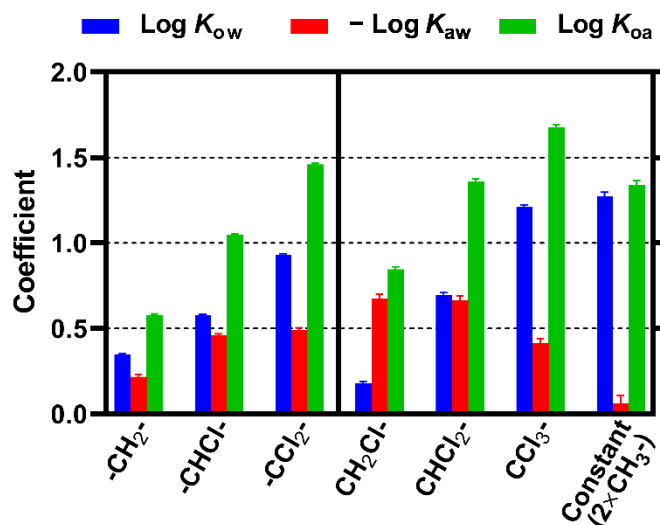


Figure S4. Fragment contributions to log  $K$ 's in the Level 1 model. In the plot, coefficients for *negative* log  $K_{aw}$  (thus log  $K_{wa}$ ) are shown for a comparison purpose. Note that the fitting constant includes the contributions of two CH<sub>3</sub>- units and that the coefficients for the substituted methyl groups indicate the contributions of replacing CH<sub>3</sub>- with these groups. This is because CPs always have two methyl groups and the number of CH<sub>3</sub>- is linearly dependent on the numbers of the other methyl groups. In contrast, the coefficients of the (substituted) methylene groups indicate the contributions of adding these groups to the molecule. Error bars indicate standard errors.

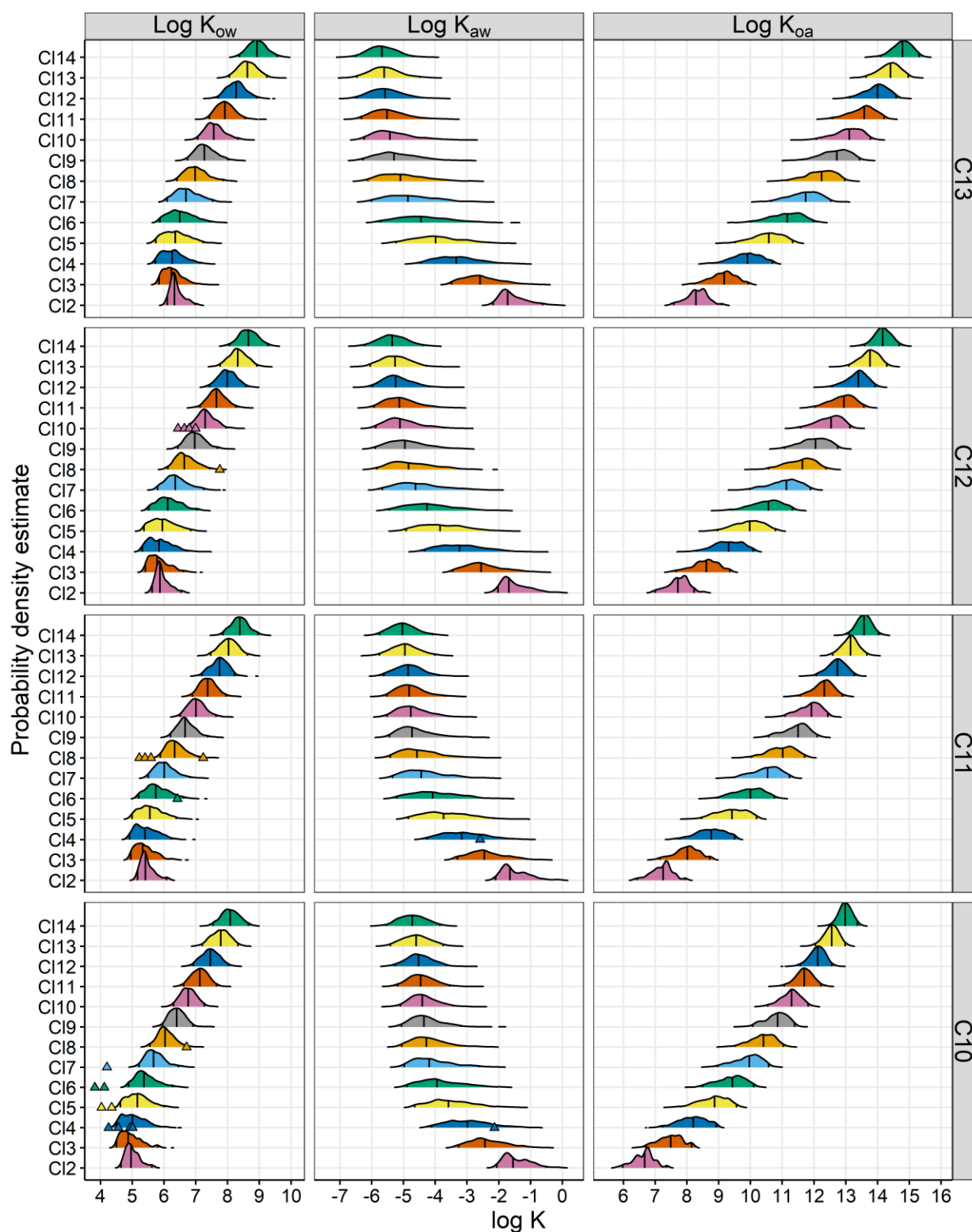


Figure S5. Kernel density estimates resulting from 1000 isomer structures for each congener group (predicted by PLSR model, with double and triple Cl substitutions allowed). Vertical lines indicate the 2.5, 50, and 97.5%iles. Data points are experimental data from Hilger et al.<sup>1</sup> and Drouillard et al.<sup>2</sup> for specific isomers.

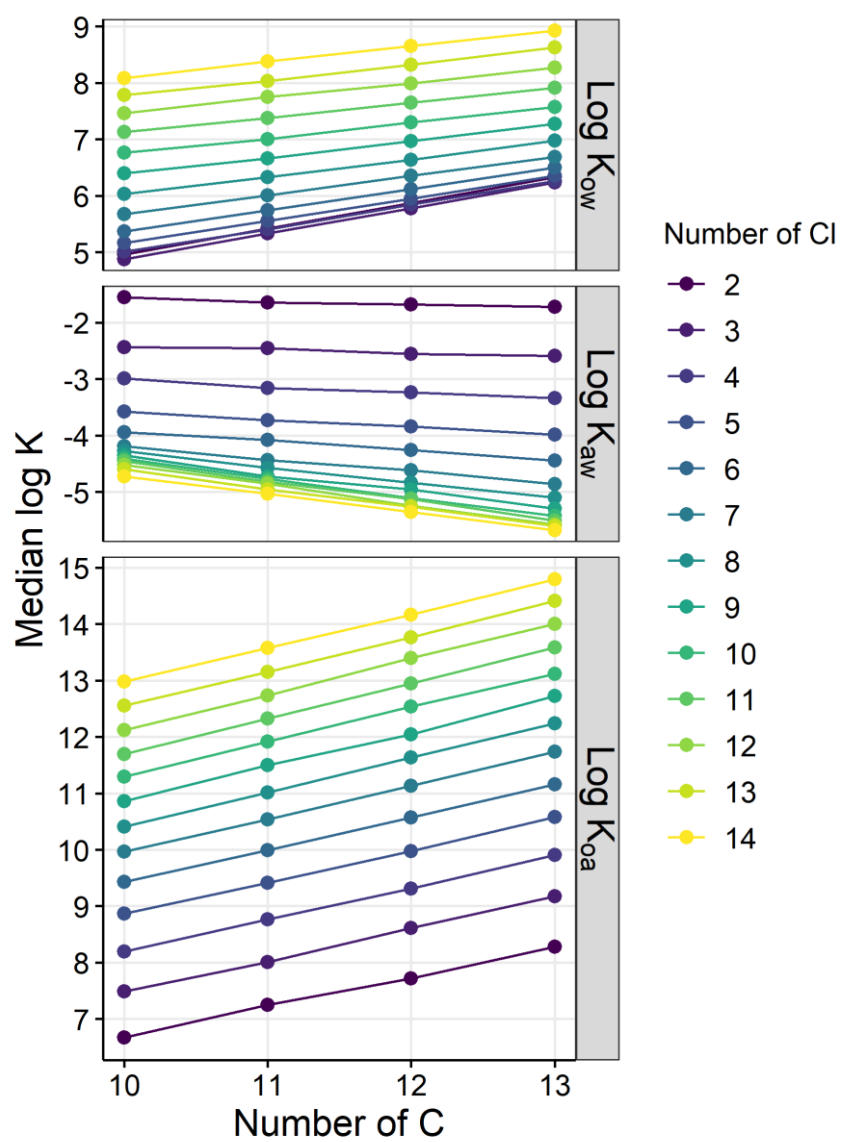


Figure S6. Dependence of the median log  $K$ 's for SCCP congener groups on the numbers of C.



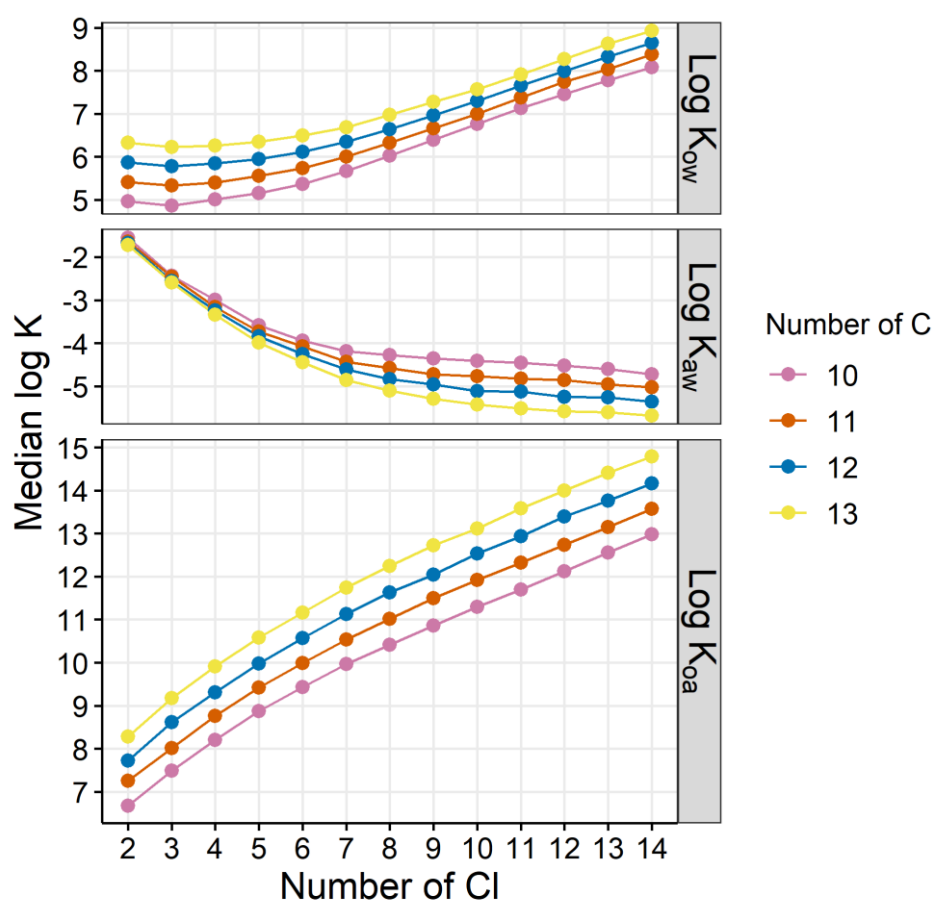


Figure S7. Dependence of the median log  $K'$ 's for SCCP congener groups on the numbers of Cl.

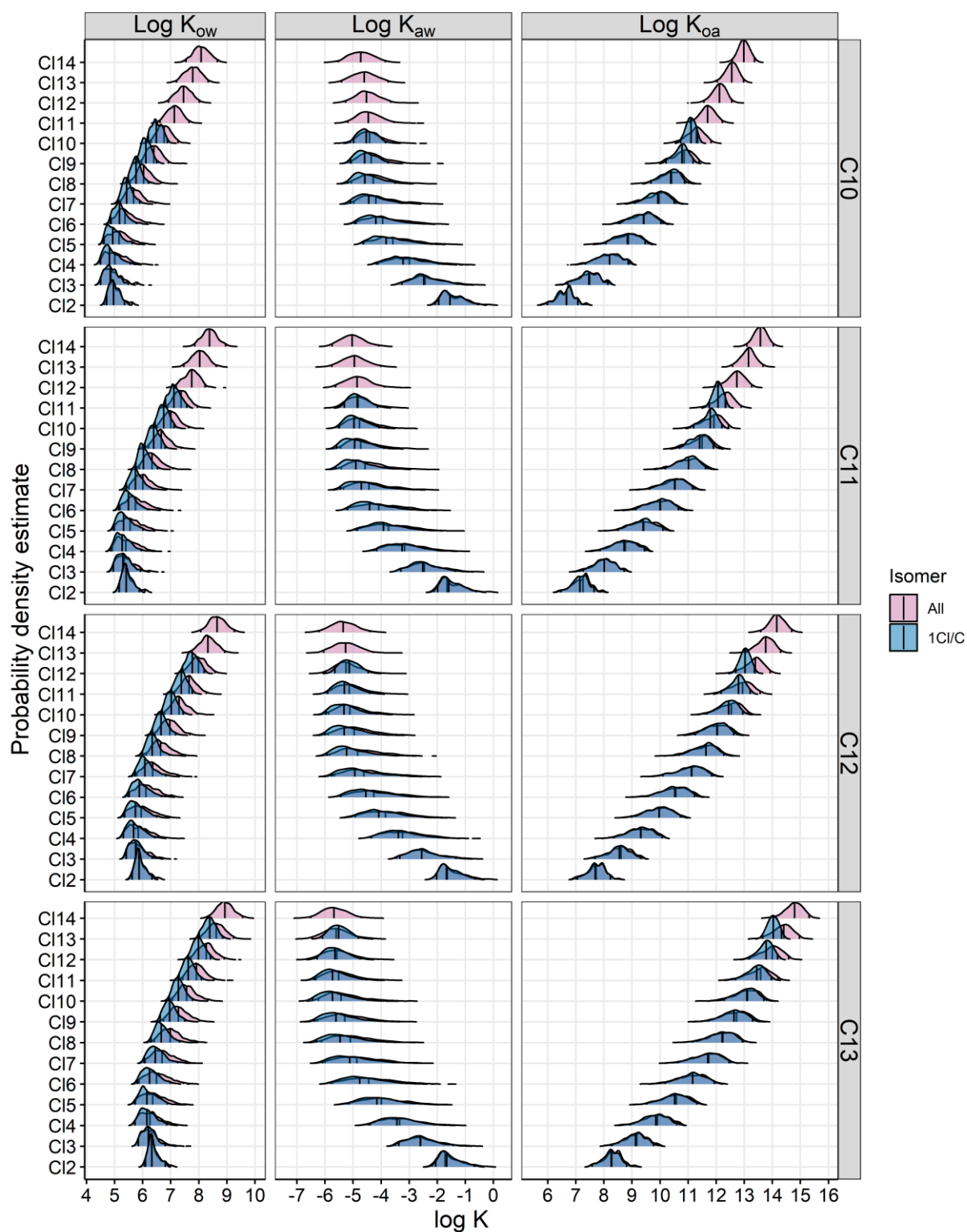


Figure S8. Kernel density estimates resulting from 1000 isomer structures for each congener group (predicted by PLSR model). “All” isomers include those with double and triple Cl substitutions, while “1 Cl/C” isomers have a maximum of one Cl atom at each C atom (double and triple Cl NOT allowed). Vertical lines indicate the 2.5, 50, and 97.5%iles.

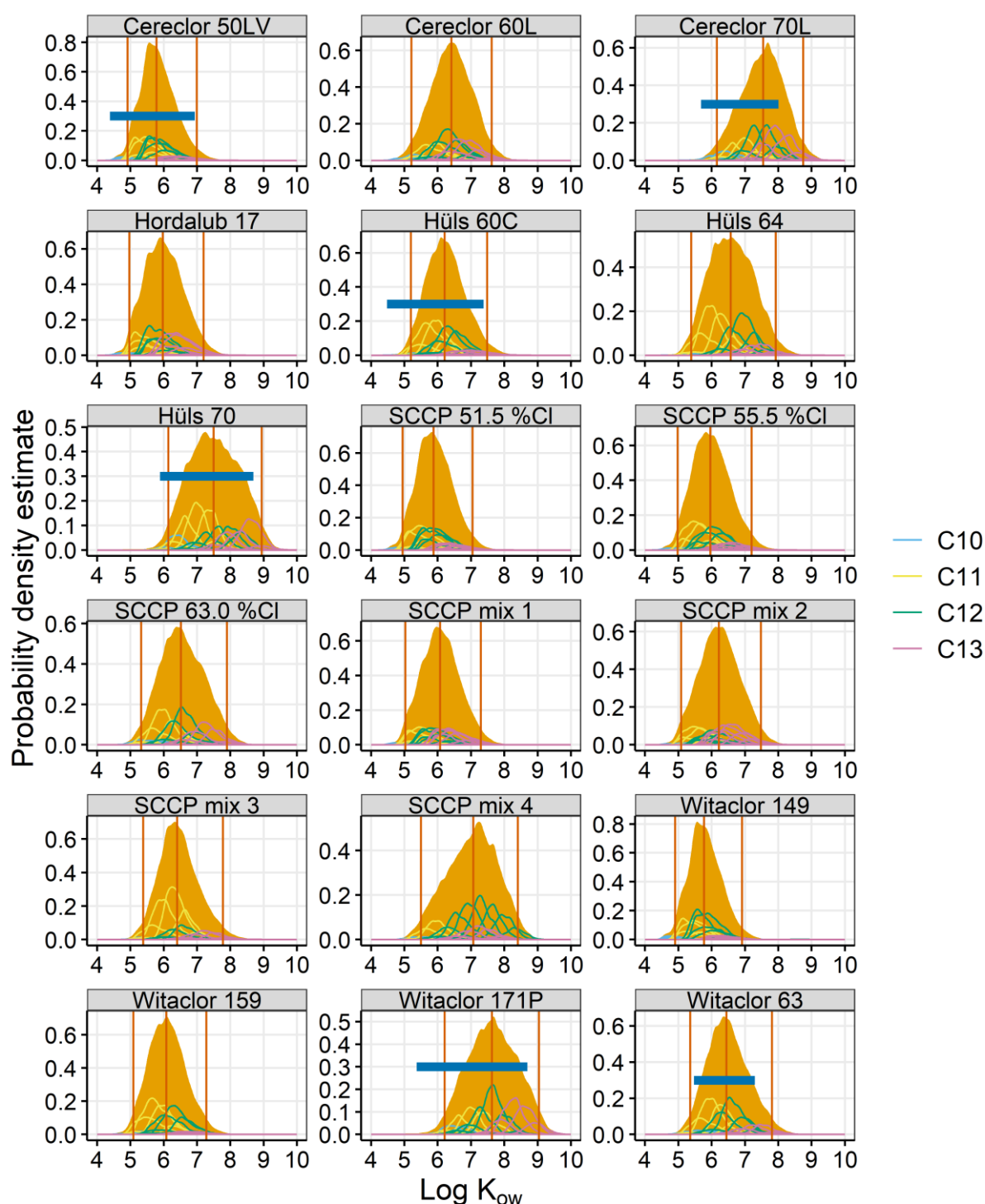


Figure S9. Distributions of log  $K_{ow}$  for CP mixtures (I). Blue bars indicate experimental data from Renberg (TLC), where available. Curves show the predictions from the FCMs (PLSR, **double/triple CI allowed**). Congener groups are displayed with different line types (e.g., dashed, dotted). SCCP mix 1, 2, 3, and 4 are, respectively, the following mixtures considered by Yuan et al.: SCCP 51.5 %CI + C<sub>13</sub> 50.23 %CI, SCCP 55.5 %CI + C<sub>13</sub> 55.03 %CI, SCCP 63.0 %CI + C<sub>11</sub> 65.25 %CI, SCCP 63.0 %CI + C<sub>12</sub> 69.98 %CI.

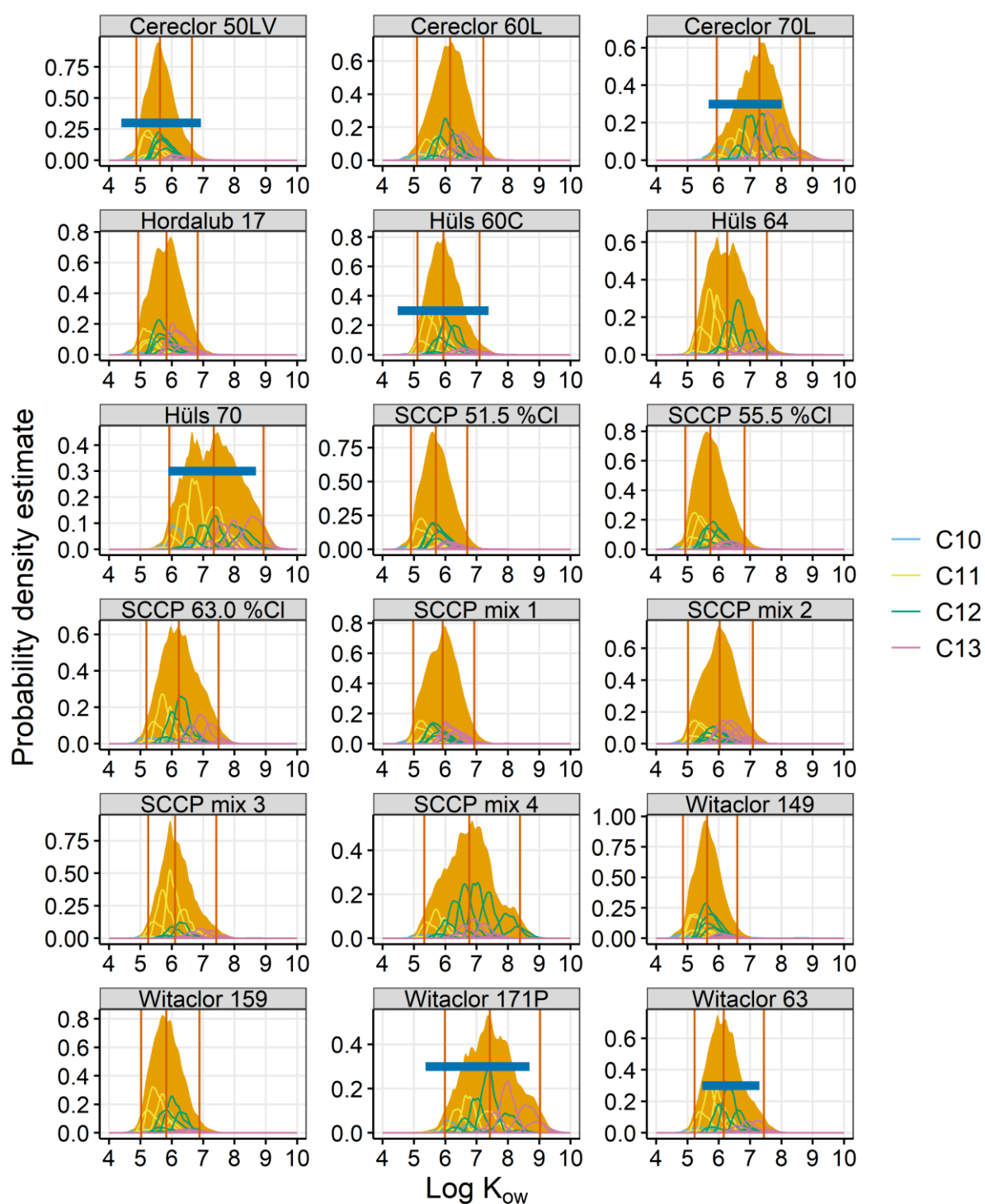


Figure S10. Distributions of  $\log K_{ow}$  for CP mixtures (II). The same plot as Figure S8, except for the predictions from the FCMs (PLSR, **double/triple CI NOT allowed**).

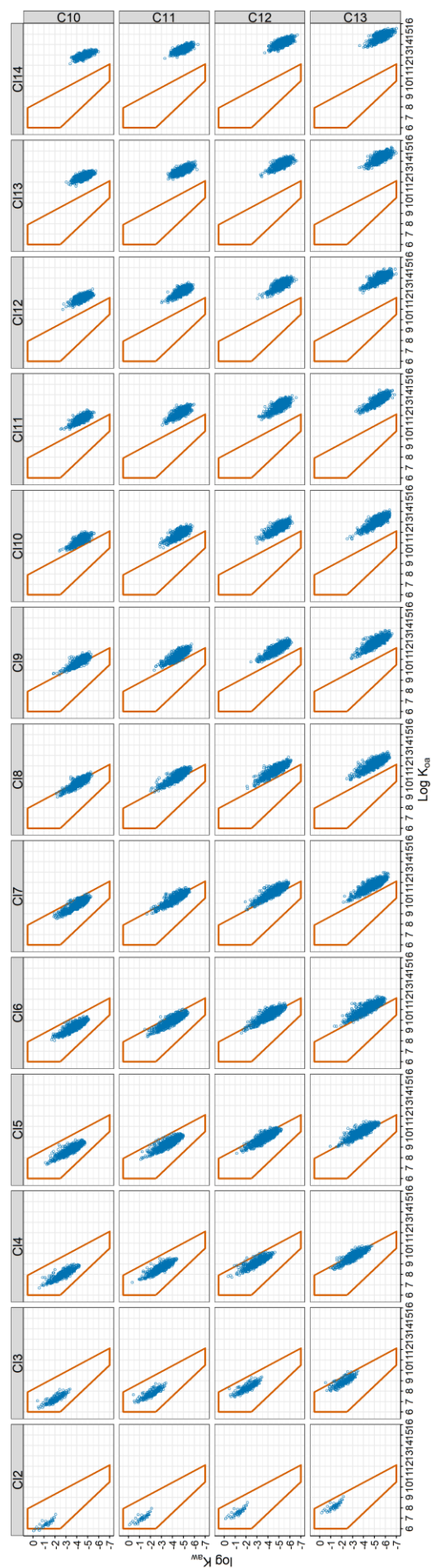


Figure S11. Chemical space plot. The chemical space for a high Arctic contamination and bioaccumulation potential AC-BAP (>10%, 70 days)<sup>3</sup> was enclosed with lines, as in ref 4.

**Table S1. Types of fragments used in the model development.**

<b>Fragment type</b>	<b>Example (terminal)</b>	<b>Example (middle)</b>	<b>Total number</b>
C <sub>1</sub>	-CH <sub>3</sub> , -CH <sub>2</sub> Cl	-CH <sub>2</sub> -, -CHCl-	7
C <sub>2</sub>	-CH <sub>2</sub> -CH <sub>3</sub> , -CCl <sub>2</sub> -CH <sub>2</sub> Cl	-CH <sub>2</sub> -CH <sub>2</sub> -, -CHCl-CH <sub>2</sub> -, -CHCl-CHCl-	19(1)
C <sub>3</sub>	-CH <sub>2</sub> -CH <sub>2</sub> -CH <sub>3</sub> , -CHCl-CCl <sub>2</sub> -CH <sub>2</sub> Cl	-CH <sub>2</sub> -CH <sub>2</sub> -CH <sub>2</sub> -, -CHCl-CHCl-CCl <sub>2</sub> -	64(10)
C <sub>4</sub>	-CH <sub>2</sub> -CH <sub>2</sub> -CH <sub>2</sub> -CH <sub>3</sub> , -CHCl-CCl <sub>2</sub> -CHCl-CCl <sub>3</sub>	-CH <sub>2</sub> -CH <sub>2</sub> -CH <sub>2</sub> -CH <sub>2</sub> -, -CHCl-CH <sub>2</sub> -CHCl-CCl <sub>2</sub> -	220(67)
<b>Model</b>	<b>Fragments used</b>		
Level 1	C <sub>1</sub>		
Level 2	C <sub>1</sub> , C <sub>2</sub>		
Level 3	C <sub>1</sub> , C <sub>2</sub> , C <sub>3</sub>		
Level 4	C <sub>1</sub> , C <sub>2</sub> , C <sub>3</sub> , C <sub>4</sub>		
PLSR	(selected fragments for the Level 4 model)		

The number in parentheses indicates the number of fragments that describe diastereomers.

Tables S2 is provided in the Excel file.

**Table S3. Results of model fitting and validation.**

Training ( $n = 815$ )				Validation ( $n = 120$ )		
	$R^2$	RMSE	AIC	Number of variables	$R^2$	RMSE
log $K_{ow}$						
Level 1	0.9833	0.183	-442	6	0.9441	0.266
Level 2	0.9879	0.155	-694	13	0.9629	0.216
Level 3	0.9965	0.083	-1641	46	0.9828	0.148
Level 4	0.9987	0.051	-2278	123	0.9881	0.123
PLSR	0.9986	0.053	*	123(22)	0.9876	0.125
log $K_{aw}$						
Level 1	0.8597	0.361	667	5	0.6389	0.494
Level 2	0.9015	0.303	395	13	0.7456	0.410
Level 3	0.9548	0.205	-192	37	0.8440	0.322
Level 4	0.9843	0.121	-876	125	0.8780	0.286
PLSR	0.9837	0.123	*	125(17)	0.8775	0.287
log $K_{oa}$						
Level 1	0.9904	0.221	-134	6	0.9623	0.340
Level 2	0.9931	0.187	-386	13	0.9709	0.297
Level 3	0.9960	0.142	-788	39	0.9763	0.263
Level 4	0.9984	0.090	-1361	122	0.9853	0.207
PLSR	0.9983	0.093	*	122(16)	0.9853	0.206

$R^2$  is the coefficient of determination (not adjusted). RMSE, root mean squared error; AIC, Akaike's Information Criterion; Values in parentheses indicate the number of PLSR components selected for the final model. \* Not applicable.

Tables S4–S6 are provided in the Excel file.

**Table S7. Predicted and experimental log  $K_{ow}$  ranges for CP mixtures. Experimental data are from Renberg et al.**

Experimental data			FCM prediction (PLSR, double/triple CI allowed)					
			Percentiles					
	min	max		0.025	0.25	0.5	0.75	0.975
Cereclor 50LV	4.39	6.93	Cereclor 50LV	4.91	5.46	5.78	6.16	6.99
Cereclor 70L	5.68	8.01	Cereclor 70L	6.16	7.06	7.54	7.97	8.75
Hüls 60C	4.48	7.38	Hüls 60C	5.20	5.82	6.21	6.61	7.49
Hüls 70L	5.89	8.69	Hüls 70	6.14	6.95	7.50	8.09	8.94
Witacclor 63	5.47	7.30	Witacclor 63	5.36	6.04	6.45	6.90	7.81
Witacclor 71P	5.37	8.69	Witacclor 171P	6.21	7.10	7.63	8.18	9.04



## References

1. Hilger, B.; Fromme, H.; Volkel, W.; Coelhan, M., Effects of chain length, chlorination degree, and structure on the octanol-water partition coefficients of polychlorinated *n*-alkanes. *Environ. Sci. Technol.* 2011, 45, (7), 2842-9.
2. Drouillard, K. G.; Tomy, G. T.; Muir, D. C. G.; Friesen, K. J., Volatility of chlorinated *n*-alkanes (C-10-C-12): Vapor pressures and Henry's law constants. *Environ. Toxicol. Chem.* 1998, 17, (7), 1252-1260.
3. Czub, G.; Wania, F.; McLachlan, M. S., Combining Long-Range Transport and Bioaccumulation Considerations to Identify Potential Arctic Contaminants. *Environ. Sci. Technol.* 2008, 42, (10), 3704-3709.
4. Brown, T. N.; Wania, F., Screening Chemicals for the Potential to be Persistent Organic Pollutants: A Case Study of Arctic Contaminants. *Environ. Sci. Technol.* 2008, 42, (14), 5202-5209.

Endo\_SCCPpartitioning\_SI\_20200622.pdf (2.67 MiB)

[view on ChemRxiv](#) • [download file](#)

---

## Other files

---

Endo\_SCCPpartitioning\_SITables\_20200622.xlsx (87.61 KiB)

[view on ChemRxiv](#) • [download file](#)

---