

## Supporting Information

### **Chlorines are Not Evenly Substituted in Chlorinated Paraffins: A Predicted NMR Pattern Matching Framework for Isomeric Discrimination in Complex Contaminant Mixtures**

Bo Yuan<sup>1\*</sup>, Daniel Henryk Lysak<sup>2</sup>, Ronald Soong<sup>2</sup>, Andrew Haddad<sup>2</sup>, Arika Hisatsune<sup>2</sup>, Arvin Moser<sup>3</sup>, Sergey Golotvin<sup>3</sup>, Dimitris Argyropoulos<sup>3</sup>, Andre J. Simpson<sup>2\*</sup>, Derek C. G. Muir<sup>4\*</sup>

<sup>1</sup> Department of Environmental Science, Stockholm University, Stockholm, Sweden, SE-10691.

<sup>2</sup> Environmental NMR Centre, University of Toronto, Toronto, ON, Canada, M1C 1A4.

<sup>3</sup> ACD/Labs, 8 King Street East, Toronto, ON, Canada, M5C 1B5.

<sup>4</sup> Canada Centre for Inland Waters, Environment and Climate Change Canada, Burlington, ON, Canada, L7S 1A1.

\*Corresponding authors:

Derek C. G. Muir, Canada Centre for Inland Waters, Environment and Climate Change Canada, Burlington, ON, Canada, L7S 1A1. Tel.: 1-905-319-6921, Fax: 1-905-336-4699, Email: [derek.muir@canada.ca](mailto:derek.muir@canada.ca)

Andre J. Simpson, Environmental NMR Centre, University of Toronto, Toronto, ON, Canada, M1C 1A4. Email: [andre.simpson@utoronto.ca](mailto:andre.simpson@utoronto.ca)

Bo Yuan, Department of Environmental Science, Stockholm University, Stockholm, Sweden, SE-10691. Email: [bo.yuan@aces.su.se](mailto:bo.yuan@aces.su.se)

## Synthesis of chlorinated paraffins

Nine samples of CP mixtures were synthesized by Quimica del Cinca (Barcelona, Spain). Tetradecane chlorinated standards were prepared by chlorination of *n*-tetradecane (>98.5%, Alfa Aesar). Pentadecane chlorinated standards were prepared by chlorination of *n*-pentadecane (99.77%, Alfa Aesar). Chlorine was directly from the supply line of CP manufacturing plant of Quimica del Cinca. The chlorination was performed in 1L glass laboratory equipment. After the preparation of each product, the chlorine and hydrogen chloride were removed by displacement with nitrogen. No stabilizer was used and the products are stored in a freezer at -18°C.

## Experimental: MS

### Sample preparation and instrumental settings

The samples were diluted with toluene<sup>1</sup> to a concentration of 30-50 ng/μL. APCI-Orbitrap-HRMS (Q Exactive, Thermo Fisher Scientific, San Jose, USA) was operated in full-scan mode (*m/z* 250 - 2000) with a resolution of 120 000 FWHM. The instrumental settings<sup>2</sup> were optimized using a short-chain chlorinated paraffin mixture (SCCP, 51.5% Cl, 15 ng/μL) and a long-chain chlorinated paraffin mixture (LCCP, Witaclor 549, 49% Cl, 5 ng/μL) as follows: injection volume 3 μL, mobile phase (acetone) flow rate 0.100 mL/min, the post-column DCM flow rate 0.010 mL/min, capillary temperature 250 °C, auxiliary gas heater temperature 250 °C, spray current 5.7 μA, maximum IT (ion time) 250 ms, automatic gain control (AGC) target of 5e6, sheath gas flow rate 17 arbs, and auxiliary gas flow rate 1 arb.

## Experimental: NMR

### Sample preparation

NMR samples were prepared by mixing 1 mL of sample in 500 μL of CDCl<sub>3</sub> (99.8 % purity, SigmaAldrich, Oakview, ON). The samples were vortexed until the sample and solvent were completely mixed and 500 μL of the mixture was transferred to a 5 mm (OD) NMR tube (Norell, Morganton, NC). The CDCl<sub>3</sub> was completely miscible with the samples given rise to a uniform and fully dissolved solution. Note the samples were quite viscous and when analyzed neat (i.e. only a trace of CDCl<sub>3</sub> for lock) the NMR line shape was broadened. It was found using 500 μL of CDCl<sub>3</sub> was enough to reduce the viscosity to permit improved line shape while still retaining a high percentage of sample required for direct <sup>13</sup>C NMR.

### NMR spectroscopy

All NMR experiments were performed on a Bruker Advance III HD spectrometer operating at a <sup>1</sup>H frequency of 500.30 MHz and equipped with a 5 mm 4 channel (<sup>1</sup>H, <sup>13</sup>C, <sup>15</sup>N, <sup>2</sup>H) Prodigy CryoProbe fitted with an actively shield Z-gradient. All samples were maintained at a temperature of 25°C during NMR data acquisition, and all channels were tuned and matched for each sample.

#### 1D <sup>1</sup>H NMR

The <sup>1</sup>H 1D NMR spectra were collected using a single 90° pulse, calibrated on a per sample basis (~ 8.5 μs) Data were collected with 32k time domain points and 15 ppm spectral width. The frequency offset was set to 4.7 ppm. For each <sup>1</sup>H NMR spectrum, 16 transients were

collected with a recycle delay of 30 s. Spectra were processed with an exponential multiplication equivalent to a 0.3 Hz line broadening with a zero-filling factor of 2 before Fourier transformation.

### *1D <sup>13</sup>C NMR*

The <sup>13</sup>C 1D NMR spectra were collected using a single 90° pulse of 12 μs with <sup>1</sup>H decoupling during acquisition. A Waltz-64 <sup>1</sup>H decoupling scheme was used with a <sup>1</sup>H B<sub>1</sub> (radio frequency) field strength of 6.25 kHz. Each <sup>13</sup>C NMR spectrum was collected using 130k time domain points and 200 ppm spectral width. The <sup>13</sup>C frequency offset was set to 100 ppm and <sup>1</sup>H frequency offset was set to 4.7 ppm.

As the <sup>13</sup>C data were used for relative quantification of the functional groups present, the <sup>13</sup>C spin lattice relaxation time (T<sub>1</sub>) were determined using the standard T<sub>1</sub> inversion recovery with Waltz-64 <sup>1</sup>H decoupling scheme. The general T<sub>1</sub> of the sample was calculated based on the recovery delay closest to the null point for <sup>13</sup>C sample resonances with longest T<sub>1</sub> values. The T<sub>1</sub> of the sample was approximated using this general formula.

$$T_{\text{null}} = \ln(2) * T_1$$

5120 transients were collected for all 1D <sup>13</sup>C NMR spectra with a recycle delay of 5\*T<sub>1</sub> (17.1s between pulses) to allow for full relaxation. Spectra were processed with an exponential multiplication equivalent to a 1 Hz line broadening in the transformed spectra and a zero-filling factor of 2.

### *2D <sup>13</sup>C-<sup>1</sup>H Heteronuclear Single Quantum Coherence (HSQC)*

The 2D HSQC (<sup>1</sup>H-<sup>13</sup>C) spectrum was collected using the Bruker standard pulse sequence (hsqcetgppsp.2) composed of two <sup>13</sup>C adiabatic chirp pulses for inversion (500 μs) and refocusing (2 ms) with GARP4 <sup>13</sup>C decoupling. Typical acquisition parameters for HSQC were: 1) <sup>1</sup>H B<sub>1</sub> field strength of 29 kHz, 2) <sup>13</sup>C B<sub>1</sub> field strength of 20 kHz, 3) a <sup>1</sup>H spectral width of 8.5 ppm with <sup>1</sup>H frequency offset equal to 4.7 ppm, 4) a <sup>13</sup>C spectral width of 160 ppm with <sup>13</sup>C frequency offset equal to 70 ppm, 5) an acquisition length of 4096 time domain points in the (F2) direct <sup>1</sup>H dimension, 6) 512 (t<sub>1</sub>) increments were collected to construct a phase sensitive (F1) <sup>13</sup>C indirect dimension via the echo/anti-echo acquisition scheme 7) 32 transients were collected for each t<sub>1</sub> increment, 8) a recycle delay of 1.5 s was used, and 9) a 3.5 kHz B<sub>1</sub> field strength was used for <sup>13</sup>C decoupling during acquisition.

The HSQC spectrum was processed via 2D Fourier transformation using an exponential multiplication equivalent to 8 Hz line broadening in the direct (F2) dimension and a sine squared function phase shifted by π/2 in the indirect (F1) dimension with a zero-filling factor of 2.

### *2D <sup>1</sup>H-<sup>13</sup>C Heteronuclear Multiple Bond Coherence (HMBC)*

The 2D HMBC (<sup>1</sup>H-<sup>13</sup>C) was collected using a Bruker standard pulse sequence (hmbcetgpl3nd)<sup>3</sup> composed of a 2 ms <sup>13</sup>C adiabatic chirp pulse for refocusing. Typical acquisition parameters for HMBC were: 1) an 8 Hz long range <sup>1</sup>H-<sup>13</sup>C J-couplings, 2) <sup>1</sup>H B<sub>1</sub> field strength of 29 kHz, 3) <sup>13</sup>C B<sub>1</sub> field strength of 20 kHz, 4) a <sup>1</sup>H spectral width of 10 ppm with <sup>1</sup>H frequency offset equal to 4.7 ppm, 5) a <sup>13</sup>C spectral width of 250 ppm with <sup>13</sup>C frequency offset equal to 100 ppm, 6) an acquisition length of 4096 time domain points in the (F2) direct <sup>1</sup>H dimension, 7) 196 (t<sub>1</sub>) increments were collected to construct a phase sensitive

(F1)  $^{13}\text{C}$  indirect dimension via the echo/anti-echo acquisition scheme 8) 32 transients were collected for each  $t_1$  increment, and 9) a recycle delay of 2 s was used.

The HMBC spectrum was processed via 2D Fourier transformation using a sine squared function phase shifted by  $\pi/2$  for both direct (F2) and indirect (F1) dimensions with a zero-filling factor of 2. The spectrum was processed in magnitude mode along the F2 dimension.

## **NMR data Prediction, Database Creation, and Spectral Matching**

### **$^{13}\text{C}$ NMR databases**

The SMILES codes for chlorinated isomeric structures were generated for molecular formulae identified as major components (by mass spectrometry) using an in-house Python script. For the  $\text{C}_{14}$  series ( $\text{C}_{14}\text{H}_y\text{Cl}_z$ ) this included all possible structures from  $\text{C}_{14}\text{H}_{29}\text{Cl}_1$  through  $\text{C}_{14}\text{H}_{20}\text{Cl}_{10}$ , while for the  $\text{C}_{15}$  series ( $\text{C}_{15}\text{H}_y\text{Cl}_z$ ) structures from  $\text{C}_{15}\text{H}_{31}\text{Cl}_1$  through  $\text{C}_{15}\text{H}_{25}\text{Cl}_7$  were considered. After removing identical structures, (i.e. mirror images that were not stereoisomers) over 410,000 unique chemical structures remained. Note that MS data showed  $\text{C}_{14}\text{Cl}_{>10}$  contributed in less than 5% to the  $\text{C}_{14}$  60.14% sample and much less (<1%) to the other samples with less chlorination. However, the number of possible structural isomers increases drastically with increased degree of chlorination (Figure S6). For example, we estimate around >250,000 isomers for  $\text{C}_{14}\text{Cl}_{11}$  and around 400,000 isomers for  $\text{C}_{14}\text{Cl}_{12}$ . If these structures were included the number of compounds would have increased to well over a million, resulting in the computational time being too onerous. As such they were not included in the study.

ACD/Labs (v.2018.2.5) was then used to predict the  $^{13}\text{C}$  chemical shifts for the generated chemical structures. A line width of 0.015ppm at half height was chosen to match the linewidth in the experimental data. Spectra were calculated using the neural network algorithm and 16,384 time domain points. The process was performed using an in-house batch predictor by ACD/Labs. Approximately 5 weeks of processing time were required to complete all the structures. The results were arranged into two databases, containing the  $\text{C}_{14}$  and  $\text{C}_{15}$  structures respectfully.

The neural network algorithm integrated into ACD/Labs is based on artificial neural nets (ANN's) as described by Blinov et al.<sup>4</sup> The approach uses an internal training database of over 2,000,000 chemical shifts and produces an error of less than 1.5 ppm (carbon) when tested against 11,000 new synthesized organic compounds (over 150,000 chemical shifts) published over the 2005-2006 period. To further confirm the ability of ACD/Labs to accurately predict chemical shifts for polychlorinated alkane species, experimental and predicted data were compared. Unfortunately, it is difficult to obtain large quantity (1 mL) of pure isomers of the species in this study, but instead three multiply chlorinated molecules were randomly selected from the Aldrich NMR spectral database and ACD/Labs was used to predict the  $^1\text{H}$  and  $^{13}\text{C}$  chemical shifts (Figure S3). In each case the observed error between the experimental data and predicted data (0.1 ppm  $^1\text{H}$  and 0.57ppm  $^{13}\text{C}$ ) is less than that reported in Blinov et al.,<sup>4</sup> showing excellent abilities of ACD's artificial neural network to predict the polychlorinated alkane species studied here.

### **Spectral Matching of the $^{13}\text{C}$ NMR to generate the top 1000**

The  $^{13}\text{C}$  NMR spectra of the chlorinated mixtures were compared against the predicted databases (Figure S2a) using ACD/Labs' similarity search algorithm. The solvent was selected as a dark region (*i.e.* not considered), the range was ordered by Euclidean distance and the

results ranked by HQI (Hit Quality Index), retaining the top 1000 matches. These top 1000 matches represent the wider range of compounds prevalent in the mixtures. The similarity search takes into account both spectral intensity and peak locations and can be summarized as follows:

Each spectrum was indexed based on its chemical shifts. For 1D  $^{13}\text{C}$  NMR spectra, the selected interval of the spectral width was divided into regions of 1 ppm and each region was indexed on a 127 point scale. The integral of each indexed region was determined and compared between the experimental spectrum and each spectrum in the database. The HQI is then calculated according to the formula:

$$\text{HQI} = 1/(N*127) \sqrt{\sum_{i=1}^L \text{abs}(p_i^{\text{spec}} - p_i^{\text{DB}})^2} \quad (\text{eq.S1})$$

where L denotes the compared regions, N is the total number of indices used and  $p_i$  is the index value (from the query spectrum and the database spectrum, respectively). In this case, a better match is indicated by a smaller HQI value. More details are provided in the ACD/Spectrus Processor Help (version 2018.2.5).

## 2D $^1\text{H}$ - $^{13}\text{C}$ NMR Databases

$^1\text{H}$ - $^{13}\text{C}$  2D NMR databases provide much higher spectral dispersion than  $^{13}\text{C}$  NMR alone. For example, Hertkorn reports the resolving power of 1D  $^{13}\text{C}$  NMR to be  $\sim 30,000$  whereas 2D  $^1\text{H}$ - $^{13}\text{C}$  HSQC approaches 2,000,000.<sup>5</sup> As such, HSQC is ideal for database matching of components in complex mixtures.<sup>6, 7</sup> Unfortunately, as the 2D database prediction and generation had to be performed manually and was considerably more computationally expensive than 1D NMR prediction, it was not possible to generate 2D NMR databases for all the 410,000 compounds. Instead  $^{13}\text{C}$  was used to pre-screen the compounds to obtain the top 1000 most likely compounds for each mixture. Each list of 1000 predicted compounds (one per mixture) took  $\sim 1$  week to complete ( $\sim 9$  weeks in total). Spectra were generated using the neural network algorithm, 1024 points in each dimension and line widths of 0.02 ppm ( $^1\text{H}$ ) and 0.2 ppm ( $^{13}\text{C}$ ). For  $^1\text{H}$ - $^{13}\text{C}$  HSQC only 1 bond correlations were included, while for  $^1\text{H}$ - $^{13}\text{C}$  HMBC 1 bond correlations were suppressed and 2-3  $^1\text{H}$ - $^{13}\text{C}$  bond correlations were included. It should be noted that both HSQC and HMBC data were used for assignment (see later) but only HSQC was used for compound matching. This is due to the fact that the number and type of correlations in HMBC (commonly across 2-5 bonds) can vary depending on a range of experimental factors. Thus the exact number of correlations obtained can differ, making it not ideal for spectral pattern matching. Conversely, HSQC detects only one bond  $^1\text{H}$ - $^{13}\text{C}$  correlations, which do not change with experimental conditions. This makes HSQC ideal for spectral pattern matching, as has been previously demonstrated for complex environmental samples.<sup>6</sup>

## Matching top 100 compounds from $^1\text{H}$ - $^{13}\text{C}$ 2D NMR

Each  $^1\text{H}$ - $^{13}\text{C}$  HSQC NMR spectrum of the chlorinated mixtures was compared against the predicted databases using a mixture search (Figure S2b). All peaks in the mixture were picked via a combination of automatic picking followed by additional manual picking to ensure all discernable peaks were selected. Mixture searches were performed giving a freedom of 3 ppm in  $^{13}\text{C}$  and 0.1 ppm in  $^1\text{H}$ , to maintain consistency with the prediction accuracies reported for ACD NMR 2D predictors.<sup>8</sup> The results were ranked by HQI, while retaining the top 100 matches. The top 100 matches represent the most likely abundant compounds in the mixtures

based on NMR but may not reflect the full diversity of compounds in the mixtures. The mixture search can be summarized as follows:

The  $HQI_{\text{mixture\_search}}$  is based on the number of peaks found in the database compared to a selected region of a query spectrum. The  $HQI_{\text{mixture\_search}}$  is calculated using the formula:

$$HQI_{\text{mixture\_search}} = 1/L \sum_{k=1}^L \frac{M_k}{N_k} \quad (\text{eq.S2})$$

Here,  $L$  is the total number of peaks regions in the query spectrum,  $N_k$  is the number of peaks in the  $k$ -peak regions of the query spectrum and  $M_k$  is the number of peaks in the  $k$ -peak region of the true spectrum.

### **NMR Assignment**

The assignments of the experimental NMR spectra were performed based on peak location and intensity matches in the ACD/Labs database (1D  $^{13}\text{C}$ , HSQC and HMBC) and then cross referenced for consistency using the short range (HSQC) and long range (HMBC) J-couplings. Importantly, the assignments were performed based on the structural motifs they represent. For each structure, the chemical shift values were collected into a spreadsheet according to their fragments and examined for their commonality. The chemical shifts of these common fragments were then mapped onto the spectra for display purposes using Mathematica 11.3. This approach to assignment provides a gauge of spectral overlap and exemplifies how complex the data are.

### **Molecular Heatmaps**

Heatmaps were generated by summing up the number of times a chlorine occurred at a specific position across the database results for each mixture. The results are expressed as simple percentages, for example for a specific position of the 10% is displayed it means that in 10% of the molecules this position contained a chlorine and 90% of the time it contained a hydrogen.

**Table S1.** Comparisons of chlorination degrees and formulae distributions of CP mixtures measured using MS and NMR.

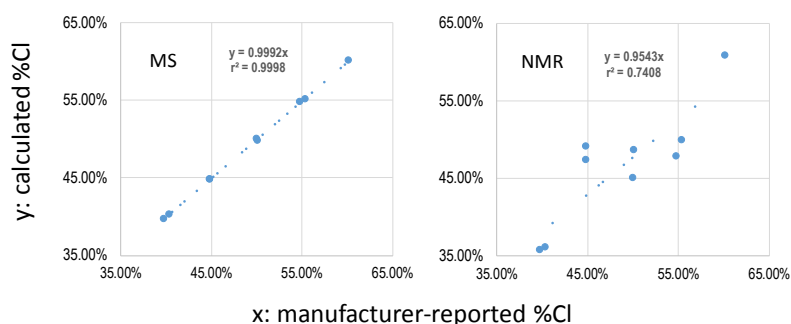
C	chlorination degree (%Cl)					the coefficient of determination (R <sup>2</sup> ) <sup>c</sup> between formulae distributions using MS and NMR (Figure S4)
	manufacturer-reported	calculated <sup>a</sup> using MS	absolute deviation <sup>b</sup> (MS)	calculated <sup>a</sup> using NMR	absolute deviation <sup>b</sup> (NMR)	
14	39.75%Cl	39.76%Cl	0.01%Cl	35.71%Cl	-4.04%Cl	0.87
	44.86%Cl	44.84%Cl	-0.02%Cl	49.03%Cl	4.17%Cl	0.59
	50.07%Cl	49.85%Cl	-0.22%Cl	48.59%Cl	-1.48%Cl	0.93
	55.34%Cl	55.15%Cl	-0.19%Cl	49.85%Cl	-5.49%Cl	0.37
	60.14%Cl	60.13%Cl	-0.01%Cl	60.79%Cl	0.65%Cl	0.85
15	40.30%Cl	40.36%Cl	0.06%Cl	36.09%Cl	-4.21%Cl	0.89
	44.80%Cl	44.82%Cl	0.02%Cl	47.28%Cl	2.48%Cl	0.83
	50.00%Cl	50.01%Cl	0.01%Cl	45.05%Cl	-4.95%Cl	0.51
	54.75%Cl	54.77%Cl	0.02%Cl	47.76%Cl	-6.99%Cl	0.30

<sup>a</sup> chlorination degree of a CP mixture was calculated by eq. S1<sup>9</sup>:

$$\%Cl = \sum[\%Cl_i \cdot \text{percentage}_i] \quad (\text{S1})$$

where %Cl<sub>*i*</sub> is the chlorination degree of molecular formula *i* (C<sub>*n*</sub>H<sub>2*n*+2-*m*</sub>Cl<sub>*m*</sub>) and percentage<sub>*i*</sub> is weight-based percentage of molecular formula *i* measured using MS.

<sup>b</sup> absolute deviation = calculated %Cl – manufacturer-reported %Cl. The correlations between two %Cl are shown below:



<sup>c</sup> R<sup>2</sup> was calculated by eq. S2<sup>10</sup>:

$$R^2 = \frac{(\sum x_i \cdot y_i - \frac{1}{n} \sum x_i \cdot \sum y_i)^2}{(\sum x_i^2 - \frac{1}{n} (\sum x_i)^2) \cdot (\sum y_i^2 - \frac{1}{n} (\sum y_i)^2)} \quad (\text{S2})$$

where *x<sub>i</sub>* and *y<sub>i</sub>* are the percentage composition of individual molecular formulae *i* in the compared two formulae distribution patterns, respectively.

**Table S2.** Major fragments and their relative quantities in the C<sub>14/15</sub> CP mixtures based on the <sup>13</sup>C NMR spectra.

Fragments*	Chemical Shift Range (ppm)	C <sub>14</sub> CP mixtures					C <sub>15</sub> CP mixtures			
		39.75%Cl	44.86%Cl	50.07%Cl	55.34%Cl	60.14%Cl	40.30%Cl	44.80%Cl	50.00%Cl	54.75%Cl
CH <sub>3</sub> -CH <sub>2</sub> -	9.7 - 20.8	9.6%	7.9%	6.0%	3.9%	2.2%	8.8%	7.4%	5.5%	3.8%
-CH <sub>2</sub> -CH <sub>2</sub> -CH <sub>2</sub> -, CH <sub>3</sub> -CH <sub>2</sub> -CH <sub>2</sub> -, CH <sub>3</sub> -CH(Cl)-	21.5 - 29.7	31.2%	24.8%	19.1%	13.9%	9.9%	31.4%	25.3%	18.7%	14.0%
C(Cl)H <sub>2</sub> -CH <sub>2</sub> -CH <sub>2</sub> -	30 - 31	5.2%	4.9%	4.9%	5.0%	4.6%	4.9%	4.6%	4.9%	5.1%
CH <sub>3</sub> -CH <sub>2</sub> -CH(Cl)-	31.9 - 32.7	1.4%	1.4%	1.5%	1.8%	1.9%	1.4%	1.5%	1.4%	1.6%
CH <sub>3</sub> -CH <sub>2</sub> -C(Cl) <sub>2</sub> -	33.2 - 36.1	7.8%	10.1%	11.5%	11.6%	8.6%	8.2%	10.5%	12.4%	12.5%
CH <sub>3</sub> -C(Cl) <sub>2</sub> -, CH(Cl) <sub>2</sub> -CH <sub>2</sub> -CH(Cl)-	36.1 - 37.5	2.6%	3.3%	3.5%	3.2%	2.4%	2.6%	3.1%	3.8%	3.5%
C(Cl)H <sub>2</sub> -CH <sub>2</sub> -CH(Cl)-, -CH <sub>2</sub> -CH(Cl)-CCl <sub>2</sub> -, CH <sub>2</sub> (Cl)-CH(Cl)-CH <sub>2</sub> -	37.5 - 40.3	11.6%	9.8%	7.5%	5.1%	4.5%	12.0%	10.4%	7.3%	4.9%
-CH(Cl)-CH <sub>2</sub> -CH(Cl), CH <sub>2</sub> (Cl)-CH <sub>2</sub> -	40.4-45.3	4.6%	5.9%	7.6%	10.0%	13.1%	4.5%	5.5%	7.3%	9.7%
CH <sub>2</sub> (Cl)-CHCl-	45.5 - 48.8	3.1%	4.2%	5.4%	6.4%	6.9%	3.2%	4.3%	5.4%	6.2%
CCl <sub>3</sub> -CH <sub>2</sub> -CH <sub>2</sub> -, -CH(Cl)-CH <sub>2</sub> -CCl <sub>2</sub> -, CH <sub>2</sub> (Cl)-CH <sub>2</sub> -	53.8 - 55.02	0.4%	0.6%	0.7%	1.0%	1.4%	0.3%	0.5%	0.8%	1.0%
-CHCl-CH <sub>2</sub> -CH <sub>2</sub> -, -CCl <sub>2</sub> -CH <sub>2</sub> -CH <sub>2</sub> -	55.5 - 56.4	0.6%	0.9%	1.6%	2.6%	3.6%	0.5%	0.9%	1.6%	2.4%
CH <sub>3</sub> -CH(Cl)-CH <sub>2</sub> -, CH <sub>3</sub> -CH <sub>2</sub> -CH(Cl)-	56.4 - 61.5	8.6%	11.2%	14.4%	18.0%	21.8%	8.6%	11.2%	14.6%	17.8%
-CH <sub>2</sub> -C(Cl)H-CH <sub>2</sub> -, -CH <sub>2</sub> -C(Cl)H- C(Cl)H-, CH <sub>3</sub> -C(Cl)H-CCl <sub>2</sub> -	61.5 - 64.2	7.7%	7.7%	7.8%	7.8%	7.9%	7.7%	7.7%	7.8%	7.8%
-CH(Cl)-CH(Cl)-CH(Cl)-, -CH <sub>2</sub> -CH(Cl)-C(Cl) <sub>2</sub> -	64.2 - 67.5	5.1%	6.0%	7.0%	7.6%	8.0%	5.0%	6.0%	6.9%	7.5%
High Chlorination	89.5 - 96.6	0.5%	0.8%	1.2%	1.9%	3.0%	0.6%	0.8%	1.2%	1.7%
Double Bond	113.2 - 134.3	0.2%	0.2%	0.2%	0.2%	0.2%	0.2%	0.2%	0.3%	0.4%
<b>Total Integral</b>		100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%

\* Some regions were combined together due to overlap, which however is as close as the values quantified directly from the 1D <sup>13</sup>C NMR data.



**Table S3.** A summary of the physicochemical properties and toxicity parameters of the C<sub>14/15</sub> CP mixtures.

C	%Cl	Log BCF	Log K <sub>OC</sub>	Log K <sub>OW</sub> <sup>a</sup>	Log S	Probability (LD <sub>50</sub> <sup>b</sup> ≤ 2000 mg/kg)	Probability (LD <sub>50</sub> <sup>b</sup> ≤ 5000 mg/kg)
14	39.75%Cl	5.10	5.20	6.97	-6.05	60.5%	90.4%
	44.86%Cl	5.04	5.15	6.79	-6.76	66.9%	92.3%
	50.07%Cl	4.87	5.03	6.53	-6.68	69.9%	93.2%
	55.34%Cl	4.78	4.97	6.57	-6.94	71.1%	93.6%
	60.14%Cl	4.99	5.11	7.21	-7.97	78.5%	95.0%
15	40.30%Cl	5.50	5.48	7.50	-6.49	55.9%	88.7%
	44.80%Cl	5.20	5.27	6.91	-6.45	64.8%	92.3%
	50.00%Cl	5.06	5.16	6.77	-6.84	70.3%	93.5%
	54.75%Cl	5.04	5.15	6.75	-6.84	70.4%	93.8%

Each parameter of a CP mixture was calculated by:

$$\text{Parameter}(\text{CP mixture}) = \sum_1^{100} \text{mole}(\text{CP isomer}_i) \times \text{parameter}(\text{CP isomer}_i) \quad (\text{eq.S3})$$

where mole(CP isomer<sub>*i*</sub>) is the molecular amount of CP isomer *i* in the CP mixture which equals one for all the isomers; parameter(CP isomer<sub>*i*</sub>) is the parameter of CP isomer *i* and is given in Worksheet (SI\_WS 1-9);

a. In the pH region where the molecule is predominantly unionized, log D=log K<sub>OW</sub><sup>11</sup>

b. Lethal Dose<sub>50</sub> is the amount of a CP mixture, given all at once, which causes the death of 50% of a group of rat/mouse.

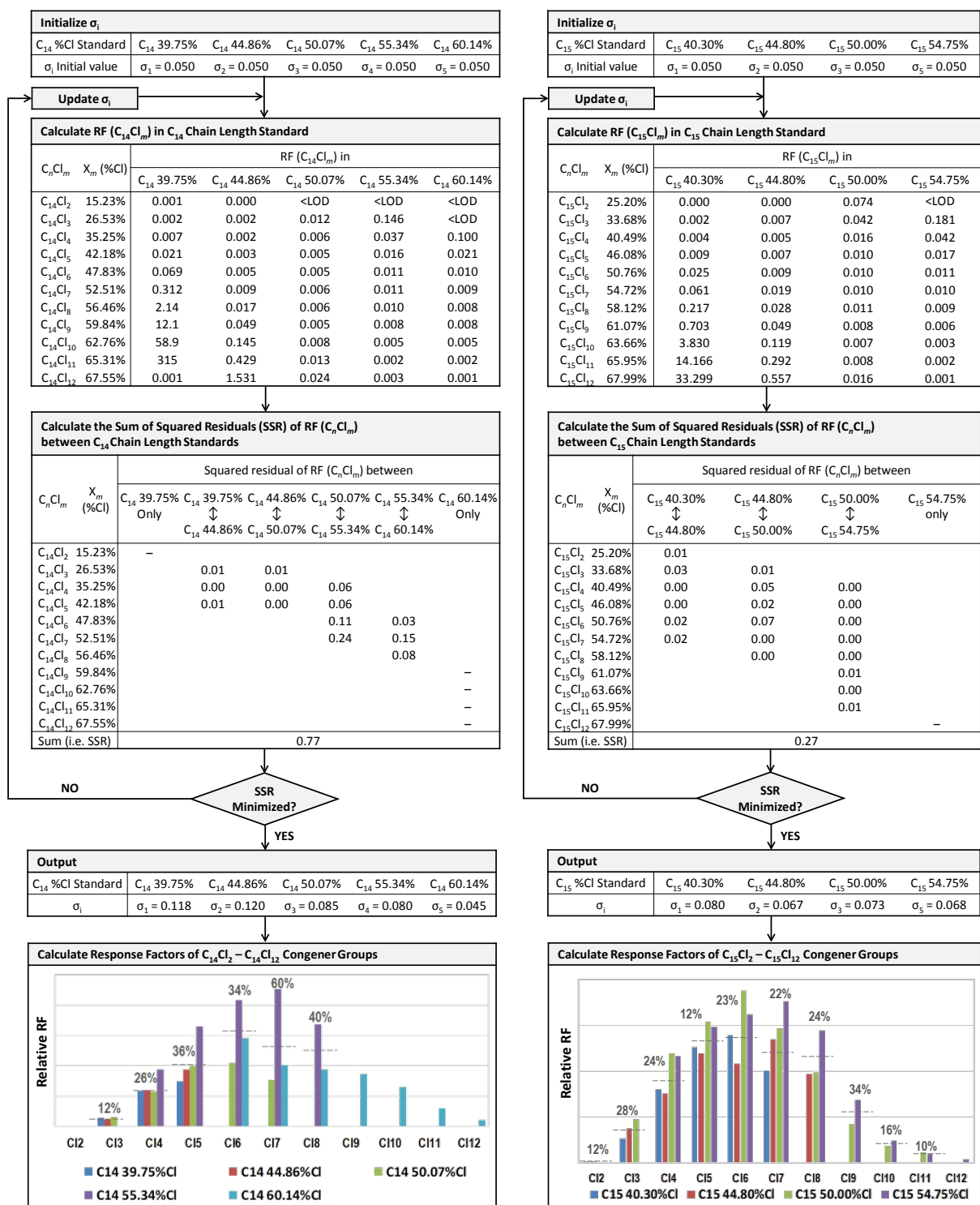
**Table S4.** Calculation of log K<sub>OW</sub> of C<sub>15</sub> 50.00% Cl CPs based on the predicted log K<sub>OW</sub> of C<sub>15</sub> congener groups in Glüge et al.<sup>12</sup>

No. of chlorine of congener group	Molar composition	Predicted log K <sub>OW</sub> <sup>a</sup>			Calculated log K <sub>OW</sub> of C <sub>15</sub> 50.00%Cl mixture
		min	max	mean <sup>b</sup>	
Cl 1	0.0%	8.22	8.29	8.26	7.78 (7.13-8.44) <sup>c</sup>
Cl 2	1.9%	–	–	–	
Cl 3	5.2%	7.37	8.30	7.84	
Cl 4	14.3%	–	–	–	
Cl 5	18.0%	7.14	8.41	7.78	
Cl 6	29.3%	–	–	–	
Cl 7	19.4%	6.99	8.48	7.74	
Cl 8	7.1%	–	–	–	
Cl 9	2.7%	7.61	8.55	8.08	
Cl10	1.0%	–	–	–	
Cl11	0.5%	6.88	8.54	7.71	
Cl12	0.5%	–	–	–	
Cl13	0.2%	7.53	8.62	8.08	

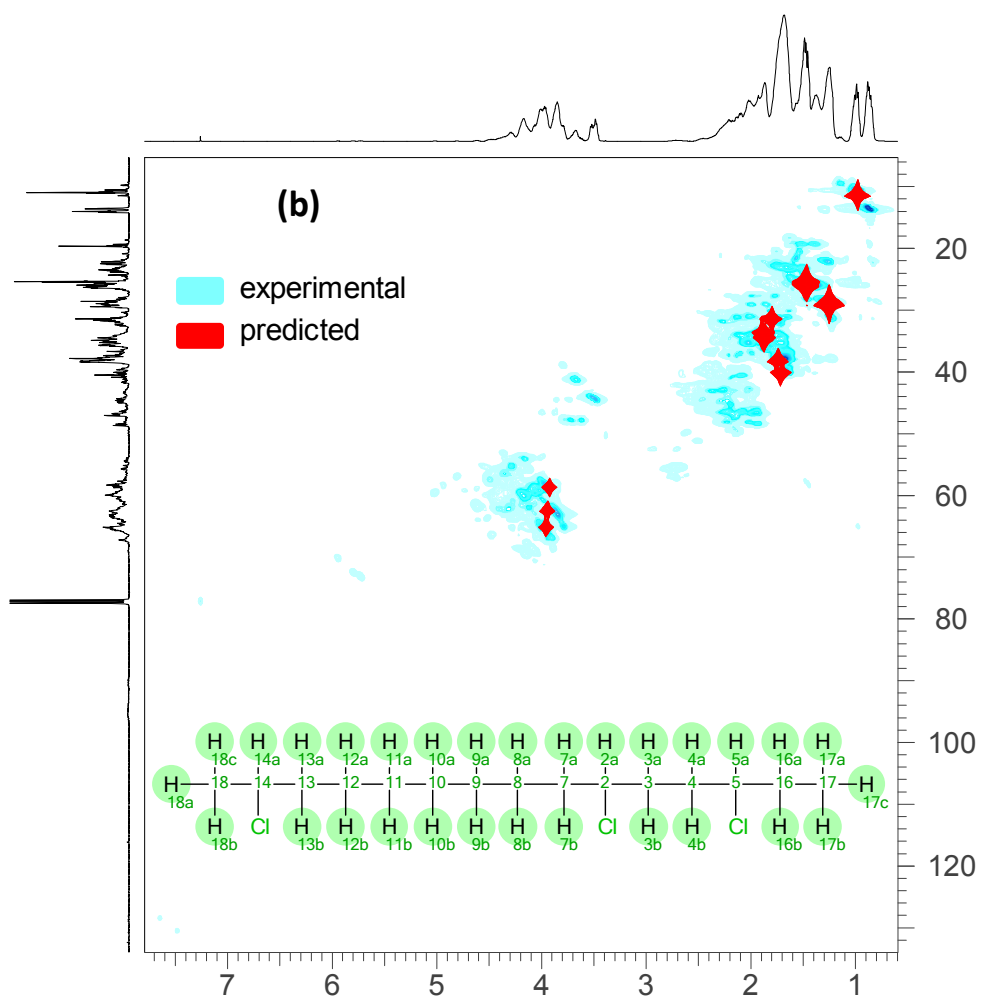
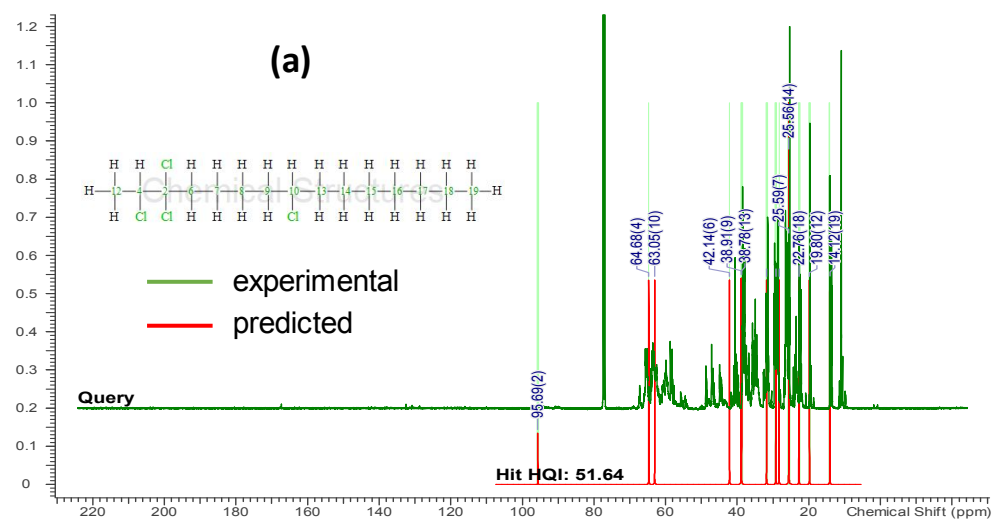
a. log K<sub>OW</sub> values of congener groups which of only odd-numbered chlorines were predicted in Glüge et al.<sup>12</sup> For each congener group, isomers with the following distribution of the chlorine atoms was assumed for prediction of log K<sub>OW</sub>: (i) at both ends of the carbon chain; (ii) in the middle of the carbon chain; (iii) equally distributed over the whole carbon chain; (iv) at one end of the carbon chain;<sup>12</sup>

b. mean of min and max values;

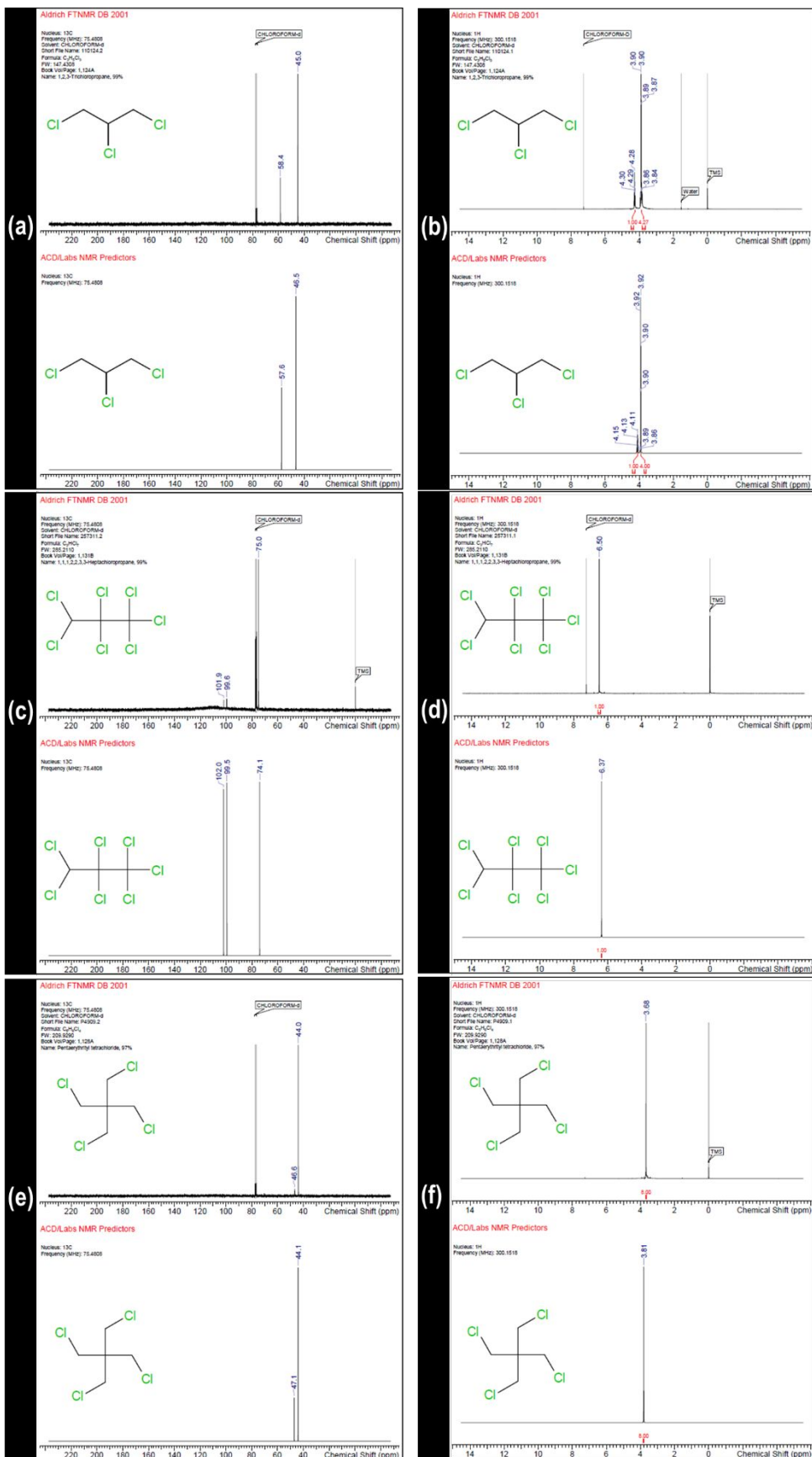
c. product mean of log K<sub>OW</sub> values of congener groups of odd-numbered chlorines; mean (min-max).



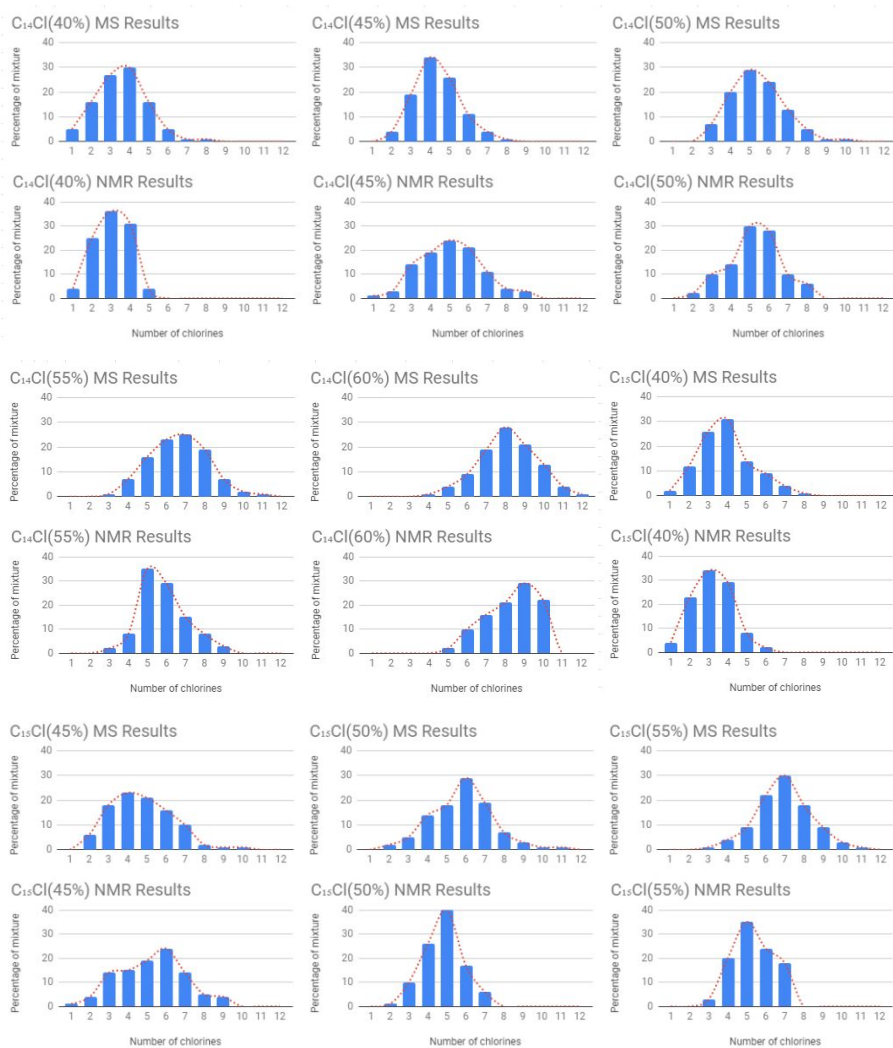
**Figure S1.** Data processing of the  $C_{14-15}$  single-chain-length standards for response factors (RFs) of individual  $C_nCl_m$  using APCI-Orbitrap-MS. The molar distribution of  $C_nCl_m$  calculated with the RFs were shown in Figure 1e and Figure S4.



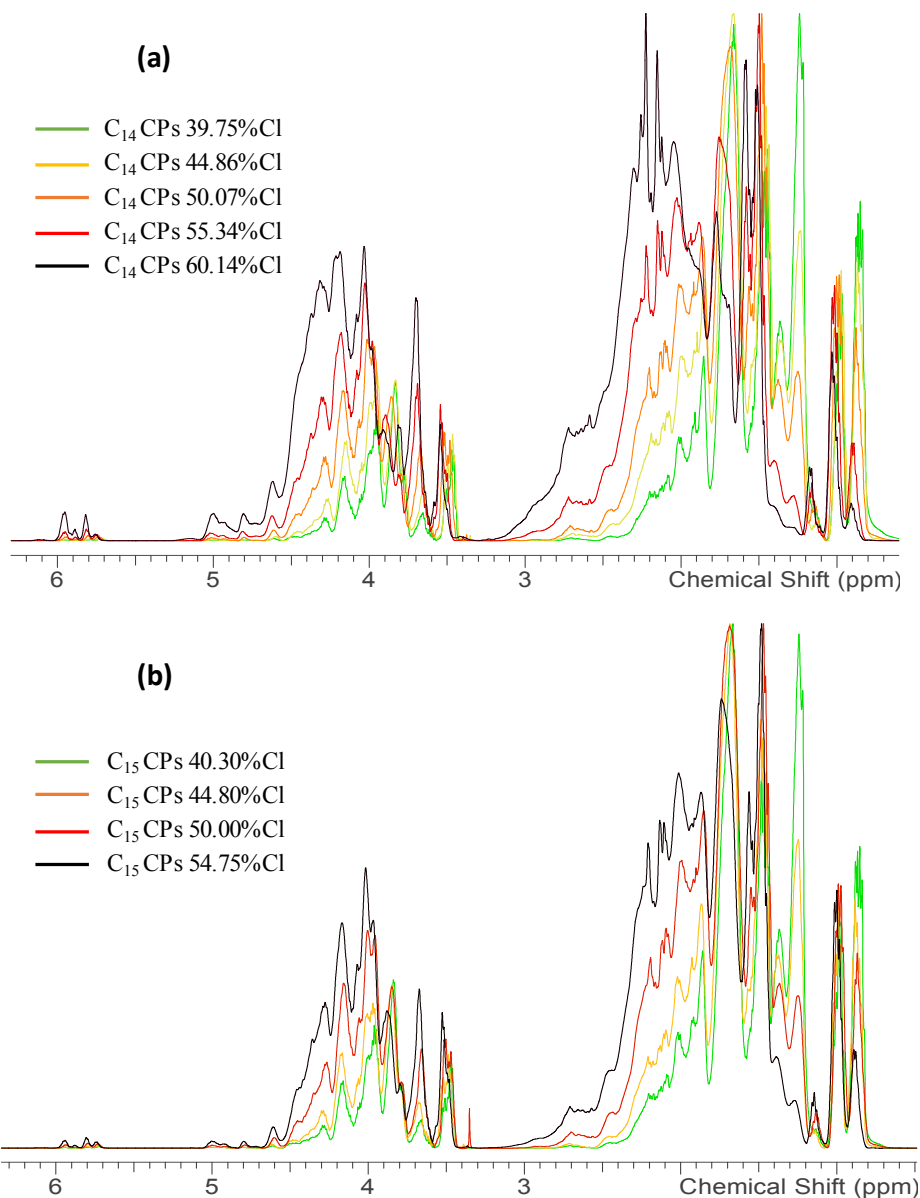
**Figure S2.** Examples of spectral matching of (a) 1D  $^{13}\text{C}$  and (b) 2D HSQC spectra to generate the top 1000 and the top 100 isomers, respectively.



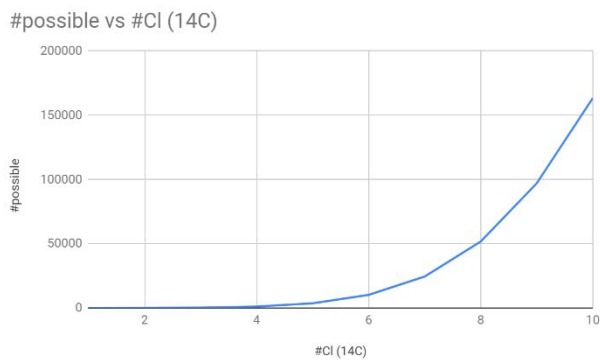
**Figure S3.** Example experimental and predicted  $^{13}\text{C}$  and  $^1\text{H}$  spectra of polychlorinated alkane species:  $^{13}\text{C}$  (a) and  $^1\text{H}$  (b) of 1,2,3-trichloropropane,  $^{13}\text{C}$  (c) and  $^1\text{H}$  (d) of 1,1,1,2,2,3,3-heptachloropropane ( $\text{C}_3\text{HCl}_7$ ), and  $^{13}\text{C}$  (e) and  $^1\text{H}$  (f) of pentaerythrityl tetrachloride.



**Figure S4.** NMR compared to MS formulae for top 100 isomers.



**Figure S5.** <sup>1</sup>H NMR spectra of five C<sub>14</sub> CP mixtures (a) and four C<sub>15</sub> CP mixtures (b).



**Figure S6.** Increasing number of possible isomers with the number of chlorines in a C<sub>14</sub> CP mixture.

## References

1. Chibwe, L.; Myers, A. L.; De Silva, A. O.; Reiner, E. J.; Jobst, K.; Muir, D.; Yuan, B., C12–30  $\alpha$ -Bromo-Chloro “Alkenes”: Characterization of a Poorly Identified Flame Retardant and Potential Environmental Implications. *Environmental Science & Technology* **2019**, *53*, (18), 10835-10844.
2. Yuan, B.; Tay, J. H.; Papadopoulou, E.; Haug, L. S.; Padilla-Sánchez, J. A.; de Wit, C. A., Complex Mixtures of Chlorinated Paraffins Found in Hand Wipes of a Norwegian Cohort. *Environmental Science & Technology Letters* **2020**, *7*, (3), 198-205.
3. Cicero, D. O.; Barbato, G.; Bazzo, R., Sensitivity Enhancement of a Two-Dimensional Experiment for the Measurement of Heteronuclear Long-Range Coupling Constants, by a New Scheme of Coherence Selection by Gradients. *Journal of Magnetic Resonance* **2001**, *148*, (1), 209-213.
4. Blinov, K. A.; Smurnyy, Y. D.; Elyashberg, M. E.; Churanova, T. S.; Kvasha, M.; Steinbeck, C.; Lefebvre, B. A.; Williams, A. J., Performance Validation of Neural Network Based <sup>13</sup>C NMR Prediction Using a Publicly Available Data Source. *Journal of Chemical Information and Modeling* **2008**, *48*, (3), 550-555.
5. Hertkorn, N.; Ruecker, C.; Meringer, M.; Gugisch, R.; Frommberger, M.; Perdue, E. M.; Witt, M.; Schmitt-Kopplin, P., High-precision frequency measurements: indispensable tools at the core of the molecular-level analysis of complex systems. *Analytical and Bioanalytical Chemistry* **2007**, *389*, (5), 1311-1327.
6. Simpson, A. J.; Lefebvre, B.; Moser, A.; Williams, A.; Larin, N.; Kvasha, M.; Kingery, W. L.; Kelleher, B., Identifying residues in natural organic matter through spectral prediction and pattern matching of 2D NMR datasets. *Magnetic Resonance in Chemistry* **2004**, *42*, (1), 14-22.
7. Tabatabaei Anaraki, M.; Dutta Majumdar, R.; Wagner, N.; Soong, R.; Kovacevic, V.; Reiner, E. J.; Bhavsar, S. P.; Ortiz Almirall, X.; Lane, D.; Simpson, M. J.; Heumann, H.; Schmidt, S.; Simpson, A. J., Development and Application of a Low-Volume Flow System for Solution-State in Vivo NMR. *Analytical Chemistry* **2018**, *90*, (13), 7912-7921.
8. Kelleher, B. P.; Simpson, A. J., Humic Substances in Soils: Are They Really Chemically Distinct? *Environmental Science & Technology* **2006**, *40*, (15), 4605-4611.
9. Yuan, B.; Bogdal, C.; Berger, U.; MacLeod, M.; Gebbink, W. A.; Alsberg, T.; de Wit, C. A., Quantifying Short-Chain Chlorinated Paraffin Congener Groups. *Environmental Science & Technology* **2017**, *51*, (18), 10633-10641.
10. Bogdal, C.; Alsberg, T.; Diefenbacher, P. S.; MacLeod, M.; Berger, U., Fast quantification of chlorinated paraffins in environmental samples by direct injection high-resolution mass spectrometry with pattern deconvolution. *Analytical Chemistry* **2015**, *87*, (5), 2852-60.
11. Hodges, G.; Eadsforth, C.; Bossuyt, B.; Bouvy, A.; Enrici, M.-H.; Geurts, M.; Kotthoff, M.; Michie, E.; Miller, D.; Müller, J.; Oetter, G.; Roberts, J.; Schowanek, D.; Sun, P.; Venzmer, J., A comparison of log K<sub>ow</sub> (n-octanol–water partition coefficient) values for non-ionic, anionic, cationic and amphoteric surfactants determined using predictions and experimental methods. *Environmental Sciences Europe* **2019**, *31*, (1), 1.
12. Gluge, J.; Bogdal, C.; Scheringer, M.; Buser, A. M.; Hungerbühler, K., Calculation of Physicochemical Properties for Short- and Medium-Chain Chlorinated Paraffins. *J Phys Chem Ref Data* **2013**, *42*, (2).