

# A comparative assessment of octanol-water partitioning and distribution constant estimation methods for perfluoroalkyl carboxylates and sulfonates

SIERRA RAYNE<sup>1\*</sup> and KAYA FOREST<sup>2</sup>

<sup>1</sup>*Ecologica Research, Penticton, British Columbia, Canada*

<sup>2</sup>*Department of Chemistry, Okanagan College, Penticton, British Columbia, Canada*

## ABSTRACT

New experimental data is available in the literature regarding the octanol-water distribution behavior of representative straight chain perfluoroalkyl carboxylate (PFCA) and sulfonate (PFSA) congeners. The current study provides the first investigation into the predictive ability of various software programs for estimating the corresponding octanol-water partitioning (log P) and distribution (log D) constants of PFCAs and PFSA. Wide predictive variation was found within and between the various methods. Several programs were able to accurately estimate the log P/D fragmental contributions of a -CF<sub>2</sub>- group for PFCAs, as well as the associated Gibbs free energies for partitioning into octanol from water due to the hydrophobic character of the perfluoroalkyl chain ( $\Delta_{\text{hydrophobic}}G_{\text{ow}}$ ). Only the SPARC log D method accurately predicted the electrostatic contributions of the carboxylate head group ( $\Delta_{\text{electrostatic}}G_{\text{ow}}$ ) towards octanol-water

partitioning for PFCAs. Similar log D values and organic carbon normalized sediment-water partitioning coefficients ( $K_{oc}$ ) for PFCAs and PFSA's having equivalent perfluoroalkyl chain lengths suggests potentially equivalent  $\Delta_{\text{electrostatic}}G_{ow}$  and  $\Delta_{\text{hydrophobic}}G_{ow}$  contributions towards lipophilic partitioning for these two contaminant classes, regardless of head group identity. In contrast, there are potentially different  $\Delta_{\text{electrostatic}}G_{ow}$  and  $\Delta_{\text{hydrophobic}}G_{ow}$  contributions towards proteinophilic partitioning.

---

\*Address correspondence to Sierra Rayne, Ecologica Research, Penticton, British Columbia, Canada; E-mail: rayne.sierra@gmail.com

**Keywords:** perfluorinated acids; perfluoroalkyl carboxylates; perfluoroalkyl sulfonates; octanol-water partitioning constant (log P); octanol-water distribution constant (log D), in silico estimation methods; Gibbs free energies; lipophilicity

## INTRODUCTION

Perfluoroalkyl carboxylic acids (PFCAs) and sulfonic acids (PFSAs) (Figure 1) are globally distributed halogenated aliphatic contaminants with ionic head groups that behave as surface active compounds in environmental and biological systems.<sup>[1-11]</sup> Because of their propensity to accumulate in abiotic and biotic matrices and resulting concerns over in vivo toxicity, their persistence towards a number of possible natural and engineered degradation processes, and their ability to achieve widespread environmental ubiquity, there is much current interest in the partitioning behavior of PFCAs and PFSAs.<sup>[11-29]</sup> Recently, the first experimental octanol-water distribution constants ( $\log D$ ) have been reported<sup>[30]</sup> for several representative straight chain PFCA and PFSA congeners (Figure 2). In light of these new experiment results, and after significant hypothetical discussion in the scientific literature over the past decade without any experimental data upon which to benchmark the debates, there is now an opportunity to assess the predictive abilities of various widely used software programs for estimating octanol-water partitioning ( $\log P$ ) and distribution ( $\log D$ ) constants of PFCAs and PFSAs. The current work thus provides the first examination of the  $\log P/D$  predictive power of various software programs towards these emerging contaminant classes.

## METHODS AND MATERIALS

PFCA and PFSA octanol-water partitioning ( $\log P_{\text{octanol:water}}$ ; hereafter referred to as  $\log P$ ) and distribution ( $\log D_{\text{octanol:water}}$ ; hereafter referred to as  $\log D$ ) constants were calculated with the

SPARC (<http://ibmlc2.chem.uga.edu/sparc/>; August 2007 release w4.0.1219-s4.0.1219) <sup>[31]</sup> and ALOGPS 2.1 (<http://www.vcclab.org/>) <sup>[32-37]</sup> software programs using the SMILES molecular formula language <sup>[38, 39]</sup> as inputs. References to PFCA and PFSA isomer patterns and congener identifications follow the framework outlined in ref. <sup>[40]</sup>. Experimental log D values for the PFCA and PFSA congeners under consideration (Figure 2) were obtained from ref. <sup>[30]</sup>.

## RESULTS AND DISCUSSION

Existing log P computational methods overpredict log D values for the straight chain C<sub>3</sub> through C<sub>9</sub> PFCA (n-PFBA through n-PFDA) and C<sub>8</sub> PFSA (n-PFOS) congeners by 1.0 to 5.4 units (Table 1). Average prediction errors within a specific computational program range from 2.5 to 4.0 units. By comparison, the ADME/Tox WEB and CLogD (using the separate VG, KLOP, and PHYS modules) log D programs generally underpredict the log D values at the experimentally determined pH range of 6 to 7 <sup>[30]</sup> by up to 4 units, with average prediction errors within a program ranging from -0.7 to -3.3 units. In contrast to the other log D prediction methods, the SPARC log D method overpredicted the log D values of both PFCAs and PFSA by between 1.0 and 3.5 units.

Wide variability among the various software programs was observed in estimating log P/D values for a particular congener, ranging between 5 to 9 units. Only the miLogP (0.34 unit error) and ClogD PHYS (-0.15 unit error) methods were able to reliably predict the log D value of n-PFOS. While the ClogD PHYS module accurately predicts the log D of n-PFOS, it performs

poorly on the PFCA congeners (errors ranging from -3.3 to -3.9 units). The ClogD VG and KLOP modules overpredict the log D of n-PFOS by 1.3 and 0.7 units, respectively, but – as noted above – underpredict the log D of the PFCAs. At pH 7, the ADME/Tox WEB log D method underpredicts the log D of both PFCAs and PFSA by between -0.4 and -1.8 units. At pH 6, however, the ADME/Tox WEB log D method accurately predicts the log D of the n-PFHpA through n-PFDA congeners (errors ranging from 0.08 to -0.2 units).

The lack of clarity regarding the experimental pH at which the PFCA log D values were determined (a range of pH 6 to 7 is quoted <sup>[30]</sup>) precludes a more definitive assessment regarding the potential log D prediction accuracy of the ADME/Tox WEB log D programs for longer chain PFCAs. For longer chain PFCAs, the pK<sub>a</sub> values may be sufficiently high (about 3 to 4 <sup>[41, 42]</sup>) such that pH dependent octanol-water partitioning may be present at pH 6 to 7. By contrast, all PFSA are expected to have pK<sub>a</sub> values <<0 <sup>[43]</sup> and thus, would not be expected to display pH dependent octanol-water partitioning behavior under environmentally or biologically relevant pH values. We have previously shown the ADME/Tox WEB pK<sub>a</sub> prediction program appears to be the only software program that accurately predicts the aqueous solution acidity constants of PFCAs.<sup>[44]</sup> For ionizable compounds such as PFCAs and PFSA, the dependence of log D values on reliable software estimation of corresponding substrate pK<sub>a</sub> values is an important determinant for obtaining reasonably accurate partitioning constant estimates. By comparison, SPARC does not adequately model the pK<sub>a</sub> values of either PFCAs or PFSA,<sup>[28, 42, 44]</sup> hindering its ability to reliably estimate corresponding log D values for these contaminants.

It is also important to reinforce the distinction between log P and log D values, particularly since log P values can be estimated for ionizing compounds. Log P values represent the relative concentrations of the molecular species in octanol and water, whereas log D values represent the relative sums of the molecular and ionized species in octanol and water.<sup>[45]</sup> Some software programs (e.g., SPARC, ADME/Tox Web, cLogP) can calculate both log P and log D. In their work, Jing et al.<sup>[30]</sup> appear to have determined the log D values for the PFCAs and PFSAAs, as they explicitly discuss measuring the concentrations of the monoanionic species in both the water and octanol phases. While we compare both estimated log D and log P values with the experimental log D values herein, we stress that a formal comparison can only be made between the experimental and estimated log D values. Any comparisons between estimated log P values and the experimental log D values are intended to examine the empirical correlation between these two values, which is an important question to consider in light of the preponderance of octanol-water partitioning constant programs that do not include substrate ionization modules. Researchers in this field need to be more clear as to whether they are calculating or measuring log P or log D values for PFCAs and PFSAAs, and include explicit caveats as to the potential theoretical inapplicability – despite any empirical correlations – of using log P values on ionizable compounds when a log D analysis may be more theoretically appropriate.

In their comparison of experimental log D and estimated log P values for PFCAs and PFSAAs, Jing et al.<sup>[30]</sup> used the >3 year old log P values from the 2006 SPARC, EPI Suite (KOWWIN), and CLogP programs as previously reported by Arp et al.<sup>[12]</sup> However, as we have shown elsewhere,<sup>[11]</sup> software programs for estimating log P/D values are continuously being refined

and updated over time. Thus, obtained values should always be considered as representing a snapshot in time, and use of historical log P/D estimates in the literature should always be confirmed with current versions of the software, particularly prior to making claims regarding accuracies of these programs. As we show in Table 1, the SPARC log P and ClogP estimates for PFCAs and PFSAAs have changed by up to 2 units since the prior work of Arp et al.,<sup>[12]</sup> and it is no longer valid to use the Arp et al.<sup>[12]</sup> dataset for these programs. In addition, Jing et al.<sup>[30]</sup> should have noted that they were comparing their experimental log D values to computationally estimated log P values.

Similarly, Kelly et al.<sup>[46]</sup> have recently not only used the 2006 SPARC log P dataset from Arp et al.<sup>[12]</sup> in their studies regarding the biological partitioning of PFCAs and PFSAAs, but these authors also applied an arbitrary and uniform one unit reduction to this out-of-date dataset in an attempt to generate purportedly reliable log P values for subsequent use in partitioning constant based classification systems. We show these final adjusted log P values from Kelly et al.<sup>[46]</sup> in Table 1, and note that they differ from the experimental log D values by a significant amount, ranging from 1.5 to 2.5 units. In addition, the errors between the log P values used by Kelly et al.<sup>[46]</sup> and the experimental log D values are not consistent with variations in chain length for the PFCAs, highlighting the errors in their arbitrary adjustment of out-of-date computational estimates by a uniform assumed error.

The availability of this straight chain series of PFCA experimental log D and estimated log P/D values with differing perfluoroalkyl chain lengths allows for the calculation of log P/D

fragmental contributions of a  $-\text{CF}_2-$  group ( $f(\text{CF}_2)$ ), as well as the associated Gibbs free energies for partitioning into octanol from water ( $\Delta G_{\text{ow}} = -RT \ln P$  or  $\Delta G_{\text{ow}} = -RT \ln D$ ) due to the electrostatic contributions of the carboxylate head group ( $\Delta_{\text{electrostatic}} G_{\text{ow}}$ ) and the hydrophobic character of the perfluoroalkyl chain ( $\Delta_{\text{hydrophobic}} G_{\text{ow}}$ ).<sup>[47]</sup> The  $f(\text{CF}_2)$  values can be obtained via linear regression of  $\log P/D$  values against the perfluoroalkyl chain length (for  $f(\text{CF}_2)$  units of  $\log P/D$ ), and via linear regression of  $\Delta G_{\text{ow}}$  against the perfluoroalkyl chain length (for  $f(\text{CF}_2)$  units of  $\text{kJ mol}^{-1}$ ). In a plot of  $\Delta G_{\text{ow}}$  against the perfluoroalkyl chain length, the y-intercept is defined as the  $\Delta_{\text{electrostatic}} G_{\text{ow}}$  value, and the difference between the value of  $\Delta G_{\text{ow}}$  for a particular compound and  $\Delta_{\text{electrostatic}} G_{\text{ow}}$  is defined as  $\Delta_{\text{hydrophobic}} G_{\text{ow}}$ . This relationship among the variables is shown in Figure 3 for the experimental PFCA  $\log D$  dataset, as well as for the 2009 SPARC  $\log P$  and  $\log D$  estimates. Where  $\log P$  values are used in such an analysis, the derived  $f(\text{CF}_2)$ ,  $\Delta_{\text{electrostatic}} G_{\text{ow}}$ , and  $\Delta_{\text{hydrophobic}} G_{\text{ow}}$  values obtained are strictly for the molecular species, whereas when  $\log D$  values are employed, the resulting  $f(\text{CF}_2)$ ,  $\Delta_{\text{electrostatic}} G_{\text{ow}}$ , and  $\Delta_{\text{hydrophobic}} G_{\text{ow}}$  values represent a composite of contributions for the molecular and ionized species in each phase. At near neutral pH values where PFSA and PFCA are effectively completely dissociated, the  $f(\text{CF}_2)$ ,  $\Delta_{\text{electrostatic}} G_{\text{ow}}$ , and  $\Delta_{\text{hydrophobic}} G_{\text{ow}}$  values obtained from  $\log D$  measurements and/or estimates are primarily representing contributions from the ionized species.

In such an analysis, it is commonly assumed that  $\Delta_{\text{electrostatic}} G_{\text{ow}}$  is constant with chain length, such that the variation in  $\Delta G_{\text{ow}}$  with changing perfluoroalkyl chain length is due solely to variation in  $\Delta_{\text{hydrophobic}} G_{\text{ow}}$ . We cannot validate this assumption, although as we have previously stressed,<sup>[11, 28, 42-44]</sup> the known substantial increase in  $\text{pK}_a$  (a potential proxy for electrostatic character of an



acidic head group) with increasing PFCA chain length ( $\text{pK}_a$  values increasing by up to about 3 units in moving from trifluoroacetic acid to n-PFOA) <sup>[41, 42]</sup> suggests that  $\Delta_{\text{electrostatic}}G_{\text{ow}}$  values may not be constant with chain length for these contaminants under some partitioning conditions. Unfortunately, the lack of experimental data and computational issues in accurately modeling the acidity of PFCAs and PFSAAs <sup>[42-44, 48]</sup> precludes any reliable conclusions as to the perfluoroalkyl chain length dependence of  $\Delta_{\text{electrostatic}}G_{\text{ow}}$  values for these two compound classes. With these possible issues in mind, we note that strong linear correlations ( $r > 0.99$ ) were observed between the experimental and estimated log P/D values for all software programs, although slight curvature in the relationship was evident for some estimates. Most importantly, the experimental dataset does not exhibit curvature in the relationship between log D and perfluoroalkyl chain length of 3 through 9 carbons, strongly suggesting that there is negligible experimental variation in  $\Delta_{\text{electrostatic}}G_{\text{ow}}$  with changing n-perfluoroalkyl chain length throughout this homologue range, and that the ionized species contributions towards  $f(\text{CF}_2)$ ,  $\Delta_{\text{electrostatic}}G_{\text{ow}}$ , and  $\Delta_{\text{hydrophobic}}G_{\text{ow}}$  values are likely dominant at the experimental pH range of 6 to 7. The lack of any changes in the electrostatic contribution towards octanol-water partitioning as a function of perfluoroalkyl chain length may also be supported by the log D finding by Jing et al. <sup>[30]</sup> for the straight chain 8:2 fluorotelomer carboxylic acid (i.e., n-PFNA with the two  $\alpha$ -fluorines next to the carboxylate head group replaced with hydrogen atoms). The 400-fold lower log D value obtained by simply replacing the two  $\alpha$ -fluorines with  $\alpha$ -hydrogens was rationalized by these authors as due to a corresponding loss of high inductive electron withdrawing power of the  $\alpha$ -fluorines, thereby increasing the ionic character of the head group and disfavoring partitioning into the less polar octanol solvent.

The experimental and estimated  $f(\text{CF}_2)$ ,  $\Delta_{\text{electrostatic}}G_{\text{ow}}$ , and  $\Delta_{\text{hydrophobic}}G_{\text{ow}}$  values for the PFCAs are given in Table 2. There is a range in the agreement between the experimental  $f(\text{CF}_2)$ ,  $\Delta_{\text{electrostatic}}G_{\text{ow}}$ , and  $\Delta_{\text{hydrophobic}}G_{\text{ow}}$  values, although several methods accurately estimate one or more of these properties. The average signed and unsigned errors between the experimental  $f(\text{CF}_2)$  value of  $-3.4 \text{ kJ mol}^{-1}$  and the various values derived from the log P/D based estimation methods are 17% and 25%, respectively. Excellent agreement was obtained with the AC\_logP, miLogP, MLOGP, CLogP (KLOP), CLogP (PHYS), CLogD (KLOP), and CLogD (PHYS) methods, having respective  $f(\text{CF}_2)$  errors of only 6%, 9%, -9%, 10%, 6%, -6%, and -8%. The  $f(\text{CF}_2)$  using 2009 data has improved for the CLogP based methods (average  $f(\text{CF}_2)$  value of  $-4.0 \text{ kJ mol}^{-1}$  and average error of 20% for the VG, KLOP, and PHYS modules), compared to the larger error (59%) reported by Ping et al. using the out-of-date 2006 CLogP data from Arp et al. The SPARC  $f(\text{CF}_2)$  error has increased to 51% using the current 2009 data compared to a 38% error in the 2006 data. No change in the KOWWIN data has occurred between 2006 and 2009. The calculation method for  $\Delta_{\text{hydrophobic}}G_{\text{ow}}$  yields an equivalent error for each method to that given for the  $f(\text{CF}_2)$  value.

Greater error is evident in the  $\Delta_{\text{electrostatic}}G_{\text{ow}}$  estimates, with average signed and unsigned errors of -41% and 107% across all software programs. All methods had absolute  $\Delta_{\text{electrostatic}}G_{\text{ow}}$  errors >75% with the exception of the SPARC log D method, which provided an accurate  $\Delta_{\text{electrostatic}}G_{\text{ow}}$  estimate within -12% of the experimental value. The ADME/Tox Web log D and CLogD methods significantly overestimated the  $\Delta_{\text{electrostatic}}G_{\text{ow}}$  value by up to 156%, whereas all log P

methods significantly underestimated the  $\Delta_{\text{electrostatic}}G_{\text{ow}}$  value by up to -164%. These error directions are consistent with the log P methods assuming that only the molecular species is dissolving in the octanol phase, which would accrue a lower free energy cost compared to the log D methods that assume the PFCAs are effectively dissociated, and therefore, that the ionized species are the dominant solutes in the octanol phase. The SPARC program only requires reparametrization of the  $f(\text{CF}_2)$  value to be able to achieve accurate log D predictions for PFCAs, whereas the AC\_logP, miLogP, MLOGP, CLogP (KLOP), CLogP (PHYS), CLogD (KLOP), and CLogD (PHYS) methods will only require reparametrization of the  $\Delta_{\text{electrostatic}}G_{\text{ow}}$  value for carboxylate ions to achieve accurate estimates. All other software programs will require reparametrization of both the  $f(\text{CF}_2)$  and  $\Delta_{\text{electrostatic}}G_{\text{ow}}$  estimation algorithms in order to achieve reliable log D estimates for PFCAs. Here we again stress that, based on first principles, log P programs should not be used to predict octanol-water partitioning constants for ionizable species such as PFCAs and PFSA, particularly if a suitable log D estimation program is available. However, given the widespread use of these log P programs on complex molecules of environmental, pharmaceutical, and industrial utility, it may be more desirable to adjust an existing log P estimation program to yield reliable log D estimates via manipulation of  $f(\text{CF}_2)$  and  $\Delta_{\text{electrostatic}}G_{\text{ow}}$  factors, rather than develop or move to an entirely new log D program.

The effective equivalence of the experimental log D values for the straight chain eight perfluorocarbon carboxylate (i.e., n-PFNA) and sulfonate (i.e., n-PFOS) suggests that the experimental  $\Delta_{\text{electrostatic}}G_{\text{ow}}$  values for perfluoroalkyl carboxylic and sulfonic acids are equivalent, as the hydrophobic contributions from the n-perfluorooctane chains on both compounds may be

equal. Caution must be exercised among researchers when comparing perfluoroalkyl chains of equivalent length. Although named n-perfluorooctanoic acid (n-PFOA), this compound only has seven perfluoroalkyl carbons in its chain, with the ‘other’ carbon being counted in the IUPAC naming system coming from the carboxylate head group. Thus, n-PFOA and n-PFOS do not have equivalent perfluoroalkyl chain lengths. While such issues appear trivial, we note that Kelly et al.<sup>[46]</sup> made this error in their work (see Table S3 in ref.<sup>[46]</sup>, whereby all PFCAs from n-PFHpA through to the n-perfluorotetradecanoic acid [n-PFTeA] have perfluoroalkyl chain lengths that are erroneously high by one carbon), leading to an incorrect relationship (both conceptual and quantitative) between perfluoroalkyl chain length and the wet weight trophic magnification factor (TMF<sub>ww</sub>) for PFCAs in Figure S7 from ref.<sup>[46]</sup>. If Kelly et al.<sup>[46]</sup> had used the correct perfluoroalkyl chain length values for PFCAs, they would have found that the TMF<sub>ww</sub> values for n-PFNA, n-PFOS, and n-perfluorooctane sulfonamide (n-PFOSA) (each of which has eight perfluoroalkyl carbons in the chain) were close in value to each other, suggesting that the hydrophobic character of the equivalent perfluoroalkyl chains for these three compounds is playing a major role in the trophic magnification behavior, and that the remaining smaller differences in the TMF<sub>ww</sub> values are due to variation in the head groups (although we stress that any trophic magnification insights regarding n-PFOSA relative to n-PFNA and n-PFOS are greatly complicated, potentially intractably, by the fact that n-PFOSA can likely be metabolized in vivo to n-PFOS).

We also note that differences in the electron withdrawing nature of the sulfonate and carboxylate groups (sulfonic acids are much stronger acids, by up to 10 pK<sub>a</sub> units or greater, than carboxylic

acids)<sup>[42, 43]</sup> may differentially alter the hydrophobic nature of the perfluoroalkyl chains for both the molecular and ionized species, leading to the experimental result whereby  $\Delta G_{ow}$  is equal for both compounds, but the composite  $\Delta_{electrostatic}G_{ow}$  and  $\Delta_{hydrophobic}G_{ow}$  values may be equally different with opposite magnitudes that offset each other to yield equivalent  $\Delta G_{ow}$  values. If the  $\Delta_{electrostatic}G_{ow}$  and  $\Delta_{hydrophobic}G_{ow}$  values for a PFCA and PFSA with the same perfluoroalkyl chain length are equal, however, this suggests that the log D values of PFCAs and PFSAs with the same chain length and branching patterns may also be effectively equal.

Even though Jing et al.<sup>[30]</sup> only determined the log D value for a single PFSA congener (n-PFOS), support for this log D equivalence hypothesis is also available from the organic carbon normalized sediment partitioning coefficient ( $K_{oc}$ ) data of Higgins and Luthy.<sup>[17, 18]</sup> In their sediment dataset,<sup>[17]</sup> the authors found very close  $K_{oc}$  values for n-PFOS and n-PFNA ( $2.57 \pm 0.13$  and  $2.39 \pm 0.09$ , respectively) and for n-perfluorodecane sulfonic acid (n-PFDS) and n-perfluoroundecane carboxylic acid (n-PFUnA) ( $3.53 \pm 0.12$  and  $3.30 \pm 0.11$ ), as well as an estimated  $f(CF_2)$  range of 0.5 to 0.6 that agrees well with the 0.61 value from Jing et al.<sup>[30]</sup> To better understand the possible close octanol-water partitioning relationship between the log D values of PFCAs and PFSAs at near neutral pH values where all compounds are expected to be dominantly in their ionized forms, additional experimental work is required using straight chain PFSAs with similar chain lengths to that examined by Jing et al.,<sup>[30]</sup> as well as on branched PFCAs and PFSAs to see if this apparent close  $\Delta G_{ow}$  relationship between straight chain members of these two classes of perfluorinated compounds also extends universally across all perfluoroalkyl chain configurations. For biological systems, the PFCAs and PFSA also behave as

proteinophilic contaminants, and bioconcentration factor (BCF) studies have generally shown that for an equivalent perfluoroalkyl chain length, PFSAAs have higher BCF values than PFCAs.

[26, 49-51] As a result, it appears that PFSAAs may have more favorable (i.e., more negative)

$\Delta_{\text{electrostatic}}G_{\text{ow}}$  values for proteinophilic partitioning than do the PFCAs, even though these two classes appear to have nearly equivalent  $\Delta_{\text{electrostatic}}G_{\text{ow}}$  values for lipophilic partitioning. While the  $\Delta_{\text{hydrophobic}}G_{\text{ow}}$  values will likely change for both PFCAs and PFSAAs in moving from a lipophilic to a proteinophilic environment, the equivalence of the perfluoroalkyl moieties from which the hydrophobic influence is derived will mean that the relative  $\Delta_{\text{hydrophobic}}G_{\text{ow}}$  values will likely remain equivalent in the new solvent system, and any net changes in the overall partitioning behavior will be driven primarily from differential changes in the  $\Delta_{\text{electrostatic}}G_{\text{ow}}$  values at the respective head groups between these two classes.

## FIGURE CAPTIONS

**Figure 1.** General structures of the compounds under study.

**Figure 2.** Structures of the PFCA and PFSA congeners for which experimental  $K_{ow}$  values are available.

**Figure 3.** Comparison between available experimental log D values for the  $C_3$  (n-PFBA) through  $C_9$  (n-PFDA) PFCA congeners (circles) and corresponding predicted values using the current versions of the SPARC log P (squares) and SPARC log D (triangles) software programs. Best fit linear regressions extrapolated negatively to the y-axis intercept are shown for each dataset. A conceptual calculation of  $\Delta_{electrostatic}G_{ow}$  and  $\Delta_{hydrophobic}G_{ow}$  values for the experimental log D dataset is also shown.

**Table 1.** Comparison of experimental and predicted log P/D values using various software programs for the straight chain C<sub>3</sub> (n-PFBA) through C<sub>9</sub> (n-PFDA) perfluoroalkyl carboxylic acids and n-perfluorooctane sulfonic acid (n-PFOS). Experimental log D errors are those given in ref. <sup>[30]</sup>. Previous log P estimates from Arp et al. <sup>[12]</sup> as cited by Jing et al. <sup>[30]</sup> and those provided by Kelly et al. <sup>[46]</sup> are also given for comparison.

	n-PFBA	n-PFPA	n-PFHxA	n-PFHpA	n-PFOA	n-PFNA	n-PFDA	n-PFOS
Perfluoroalkyl chain length	3	4	5	6	7	8	9	8
Experimental log D <sup>a</sup>	-0.52	0.09	0.70±0.10	1.31	1.92	2.57±0.07	2.90±0.10	2.45±0.08
Kelly et al. <sup>[46]</sup>	n/a	n/a	n/a	2.80	3.60	4.50	5.40	4.30
SPARC 2009 log P	2.91	3.69	4.50	5.36	6.26	7.23	8.26	4.67
SPARC 2006 log P <sup>b</sup>	n/a	n/a	3.12	3.82	4.59	5.45	6.38	5.28
SPARC log D <sub>pH6-7</sub>	0.86	1.69	2.53	3.41	4.33	5.32	6.36	3.41
ALOGPS	2.57	2.93	3.37	3.93	4.29	4.55	4.85	4.24
AC_logP	1.75	2.37	2.99	3.61	4.23	4.85	5.47	3.46
AB/logP	1.72	3.63	3.88	4.52	5.17	5.81	6.45	5.00
miLogP	1.76	2.40	3.05	3.69	4.33	4.97	5.61	2.79
ALOGP	1.84	2.33	2.83	3.32	3.81	4.30	4.79	4.16
MLOGP	1.90	2.53	3.11	3.66	4.17	4.66	5.12	3.88
KOWWIN 2009 log P	2.43	3.40	4.37	5.33	6.30	7.27	8.23	6.28
KOWWIN 2006 log P <sup>b</sup>	n/a	n/a	4.37	5.33	6.30	7.27	8.23	6.28
XLOGP2	1.70	2.44	3.18	3.93	4.67	5.42	6.16	4.89
XLOGP3	2.25	2.92	3.59	4.26	4.93	5.60	6.28	5.02
ADME/Tox WEB log D <sub>pH6</sub>	-2.36	-0.46	-0.21	1.07	1.70	2.34	2.98	0.90
ADME/Tox WEB log D <sub>pH7</sub>	-2.36	-0.46	-0.21	0.55	1.20	1.83	2.47	0.90
ClogP (VG)	2.61	3.45	4.28	5.12	5.95	6.78	7.62	5.98
ClogP (KLOP)	2.33	2.98	3.62	4.27	4.91	5.56	6.20	5.62
ClogP (PHYS)	1.98	2.61	3.23	3.85	4.47	5.10	5.72	4.69
ClogP 2006 <sup>b</sup>	n/a	n/a	3.15	3.39	3.62	3.86	4.09	2.28
ClogD (VG)	-3.26	-2.94	-2.40	-1.60	-0.77	0.06	0.90	3.71
ClogD (KLOP)	-3.50	-3.31	-2.89	-2.27	-1.62	-0.98	-0.33	3.14
ClogD (PHYS)	-3.83	-3.63	-3.20	-2.60	-1.98	-1.35	-0.73	2.30

<sup>a</sup> from ref. <sup>[30]</sup>. <sup>b</sup> from ref. <sup>[12]</sup>.



**Table 2.** Experimental and estimated log P/D fragmental contributions of a -CF<sub>2</sub>- group ( $f(\text{CF}_2)$ ) and the Gibbs free energies for the electrostatic ( $\Delta_{\text{electrostatic}}G$ ) and hydrophobic ( $\Delta_{\text{hydrophobic}}G$ ) contributions towards octanol-water partitioning using various software programs for the straight chain C<sub>3</sub> (n-PFBA) through C<sub>9</sub> (n-PFDA) perfluoroalkyl carboxylic acids.

	$f(\text{CF}_2)$ (log units)	$f(\text{CF}_2)$ (kJ mol <sup>-1</sup> )	$\Delta_{\text{electrostatic}}G$ (kJ mol <sup>-1</sup> )	$\Delta_{\text{hydrophobic}}G$ (kJ mol <sup>-1</sup> )						
				n-PFBA	n-PFPA	n-PFHxA	n-PFHpA	n-PFOA	n-PFNA	n-PFDA
Experimental log D <sup>a</sup>	0.61	-3.4±0.1	12.8±0.6	-10.1	-13.4	-16.8	-20.1	-23.5	-26.8	-30.2
SPARC 2009 log P	0.89±0.02	-5.1±0.1	-0.7±0.7	-15.9	-20.3	-25.0	-29.9	-35.0	-40.6	-46.4
SPARC 2006 log P <sup>b</sup>	0.82±0.03	-4.7±0.1	5.9	n/a	n/a	-23.7	-27.7	-32.1	-37.0	-42.3
SPARC log D <sub>pH7</sub>	0.91±0.02	-5.2±0.1	11.3±0.6	-16.2	-20.9	-25.7	-30.7	-36.0	-41.6	-47.6
ALOGPS	0.39±0.02	-2.2±0.1	-8.1±0.7	-6.5	-8.6	-11.1	-14.3	-16.3	-17.8	-19.5
AC_logP	0.62±0.00	-3.5±0.0	0.6±0.0	-10.6	-14.2	-17.7	-21.2	-24.8	-28.3	-31.8
AB_logP	0.71±0.07	-4.0±0.4	-1.2±2.7	-8.7	-19.6	-21.0	-24.6	-28.4	-32.0	-35.7
miLogP	0.64±0.00	-3.7±0.0	0.9±0.0	-11.0	-14.6	-18.3	-22.0	-25.6	-29.3	-33.0
ALOGP	0.49±0.00	-2.8±0.0	-2.1±0.0	-8.4	-11.2	-14.1	-16.9	-19.7	-22.5	-25.2
MLOGP	0.54±0.01	-3.1±0.1	-2.2±0.5	-8.7	-12.3	-15.6	-18.7	-21.6	-24.4	-27.0
KOWWIN 2009 log P	0.97±0.00	-5.5±0.0	2.7±0.0	-16.5	-22.1	-27.6	-33.1	-38.6	-44.2	-49.6
KOWWIN 2006 log P <sup>b</sup>	0.97±0.00	-5.5±0.0	2.7±0.0	-16.5	-22.1	-27.6	-33.1	-38.6	-44.2	-49.6
XLOGP2	0.74±0.00	-4.2±0.0	3.1±0.0	-12.8	-17.0	-21.2	-25.5	-29.7	-34.0	-38.2
XLOGP3	0.67±0.00	-3.8±0.0	-1.3±0.0	-11.5	-15.3	-19.1	-23.0	-26.8	-30.6	-34.5
ADME/Tox WEB log D <sub>pH6</sub>	0.84±0.07	-4.8±0.4	24.7±2.7	-11.2	-22.0	-23.5	-30.8	-34.4	-38.0	-41.7
ADME/Tox WEB log D <sub>pH7</sub>	0.73±0.07	-4.2±0.4	22.6±2.6	-9.1	-20.0	-21.4	-25.7	-29.4	-33.0	-36.7
CLogP (VG)	0.83±0.00	-4.8±0.0	-0.6±0.0	-14.3	-19.1	-23.8	-28.6	-33.3	-38.1	-42.9
CLogP (KLOP)	0.65±0.00	-3.7±0.0	-2.3±0.0	-11.0	-14.7	-18.4	-22.1	-25.8	-29.5	-33.1
CLogP (PHYS)	0.62±0.00	-3.6±0.0	-0.7±0.0	-10.6	-14.2	-17.8	-21.3	-24.9	-28.5	-32.0
CLogP 2006 <sup>b</sup>	0.24±0.01	-1.4±0.0	-11.3	n/a	n/a	-6.7	-8.1	-9.4	-10.7	-12.1
CLogD (VG)	0.72±0.04	-4.1±0.2	32.8±1.5	-14.2	-16.0	-19.1	-23.6	-28.4	-33.1	-37.9
CLogD (KLOP)	0.55±0.04	-3.1±0.2	31.0±1.3	-11.1	-12.1	-14.5	-18.1	-21.8	-25.4	-29.1
CLogD (PHYS)	0.54±0.03	-3.1±0.2	32.6±1.2	-10.7	-11.8	-14.3	-17.7	-21.3	-24.9	-28.4

<sup>a</sup> from ref. <sup>[30]</sup>, <sup>b</sup> from ref. <sup>[12]</sup>.

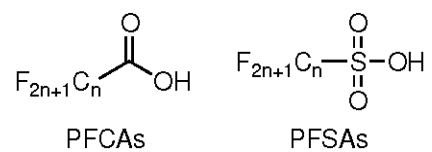


Fig. 1

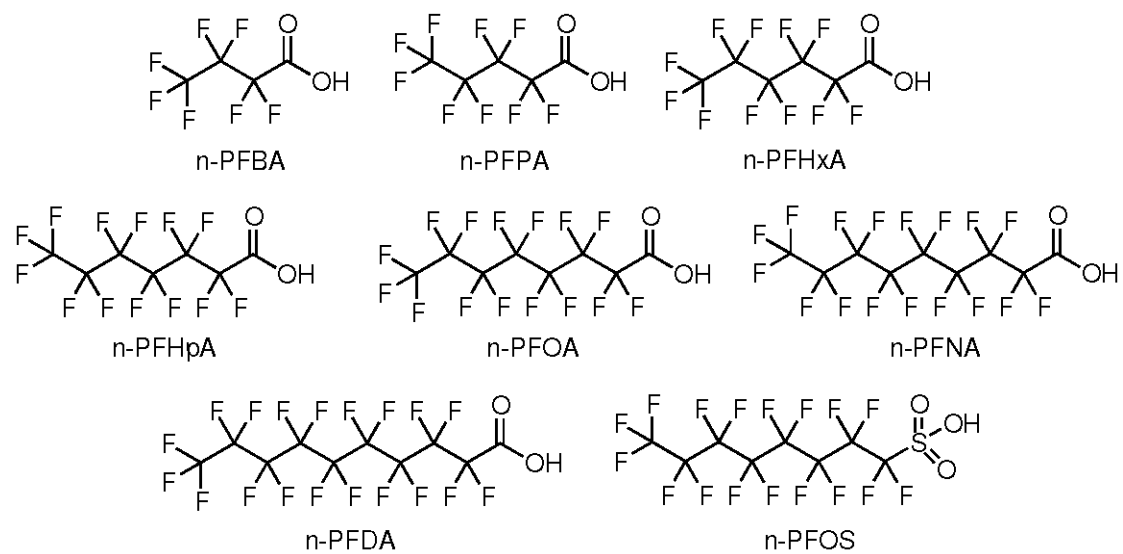


Fig. 2

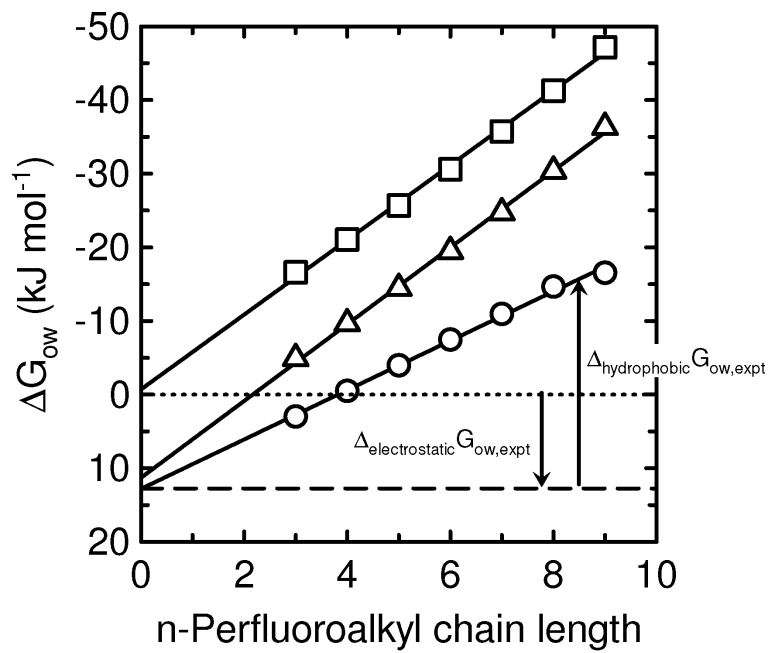


Fig. 3

## REFERENCES

- [1] Beach, S.A.; Newsted, J.L.; Coady, K.; Giesy, J.P. Ecotoxicological evaluation of perfluorooctanesulfonate (PFOS). *Rev. Environ. Contam. Toxicol.* **2006**, *186*, 133-174.
- [2] Boutonnet, J.C.; Bingham, P.; Calamari, D.; de Rooij, C.; Franklin, J.; Kawano, T.; Libre, J.; McCulloch, A.; Malinverno, G.; Odom, J.M.; Rusch, G.M.; Smythe, K.; Sobolev, I.; Thompson, R.; Tiedje, J.M. Environmental risk assessment of trifluoroacetic acid. *Human Ecol. Risk Assess.* **1999**, *5*, 59-124.
- [3] Conder, J.M.; Hoke, R.A.; De Wolf, W.; Russell, M.H.; Buck, R.C. Are PFCAs bioaccumulative? A critical review and comparison with regulatory criteria and persistent lipophilic compounds. *Environ. Sci. Technol.* **2008**, *42*, 995-1003.
- [4] Ellis, D.; Moody, C.; Mabury, S. Trifluoroacetic acid and longer chain perfluoro acids: Sources and analysis. *Organofluorines*; Neilson, A.H., Ed.; Springer: New York, NY, USA, 2002; 103-120.
- [5] Fromme, H.; Tittlemier, S.A.; Volkel, W.; Wilhelm, M.; Twardella, D. Perfluorinated compounds – Exposure assessment for the general population in western countries. *Int. J. Hyg. Environ. Health* **2009**, *212*, 239-270.
- [6] Giesy, J.P.; Kannan, K. Perfluorochemical surfactants in the environment. *Environ. Sci. Technol.* **2002**, *36*, 146A-152A.
- [7] Houde, M.; Martin, J.W.; Letcher, R.J.; Solomon, K.R.; Muir, D.C.G. Biological monitoring of polyfluoroalkyl substances: A review. *Environ. Sci. Technol.* **2006**, *40*, 3463-3473.
- [8] Lau, C.; Anitole, K.; Hodes, C.; Lai, D.; Pfahles-Hutchens, A.; Seed, J. Perfluoroalkyl acids: A review of monitoring and toxicological findings. *Toxicol. Sci.* **2007**, *99*, 366-394.

- [9] Paul, A.G.; Jones, K.C.; Sweetman, A.J. A first global production, emission, and environmental inventory for perfluorooctane sulfonate. *Environ. Sci. Technol.* **2009**, *43*, 386-392.
- [10] Prevedouros, K.; Cousins, I.T.; Buck, R.C.; Korzeniowski, S.H. Sources, fate and transport of perfluorocarboxylates. *Environ. Sci. Technol.* **2006**, *40*, 32-44.
- [11] Rayne, S.; Forest, K. Perfluoroalkyl sulfonic and carboxylic acids: A critical review of physicochemical properties, levels and patterns in waters and waste waters, and treatment methods. *J. Environ. Sci. Health A* **2009**, submitted.
- [12] Arp, H.P.H.; Niederer, C.; Goss, K. Predicting the partitioning behavior of various highly fluorinated compounds. *Environ. Sci. Technol.* **2006**, *40*, 7298-7304.
- [13] Barton, C.A.; Kaiser, M.A.; Russell, M.H. Partitioning and removal of perfluorooctanoate during rain events: The importance of physical-chemical properties. *J. Environ. Monit.* **2007**, *9*, 839-846.
- [14] Bowden, D.J.; Clegg, S.L.; Brimblecombe, P. The Henry's law constant of trifluoroacetic acid and its partitioning into liquid water in the atmosphere. *Chemosphere* **1996**, *32*, 405-420.
- [15] Ellis, D.A.; Webster, E. Response to Comment on "Aerosol enrichment of the surfactant PFO and mediation of the water-air transport of gaseous PFOA". *Environ. Sci. Technol.* **2009**, *43*, 1234-1235.
- [16] Goss, K.; Bronner, G. What is so special about the sorption behavior of highly fluorinated compounds?. *J. Phys. Chem. A* **2006**, *110*, 9518-9522.
- [17] Higgins, C.P.; Luthy, R.G. Sorption of perfluorinated surfactants on sediments. *Environ. Sci. Technol.* **2006**, *40*, 7251-7256.

- [18] Higgins, C.P.; Luthy, R.G. Modeling sorption of anionic surfactants onto sediment materials: An a priori approach for perfluoroalkyl surfactants and linear alkylbenzene sulfonates. *Environ. Sci. Technol.* **2007**, *41*, 3254-3261.
- [19] Higgins, C.P.; McLeod, P.B.; MacManus-Spencer, L.A.; Luthy, R.G. Bioaccumulation of perfluorochemicals in sediments by the aquatic oligochaete *Lumbriculus variegatus*. *Environ. Sci. Technol.* **2007**, *41*, 4600-4606.
- [20] Johnson, R.L.; Anschutz, A.J.; Smolen, J.M.; Simcik, M.F.; Penn, R.L. The adsorption of perfluorooctane sulfonate onto sand, clay, and iron oxide surfaces. *J. Chem. Engr. Data* **2007**, *52*, 1165-1170.
- [21] Jones, P.D.; Hu, W.; De Coen, W.; Newsted, J.L.; Giesy, J.P. Binding of perfluorinated fatty acids to serum proteins. *Environ. Toxicol. Chem.* **2003**, *22*, 2639-2649.
- [22] Kutsuna, S.; Hori, H. Experimental determination of Henry's law constants of trifluoroacetic acid at 278–298 K. *Atmos. Environ.* **2008**, *42*, 1399-1412.
- [23] Li, H.; Ellis, D.; Mackay, D. Measurement of low air-water partition coefficients of organic acids by evaporation from a water surface. *J. Chem. Engr. Data* **2007**, *52*, 1580-1584.
- [24] McMurdo, C.J.; Ellis, D.A.; Webster, E.; Butler, J.; Christensen, R.D.; Reid, L.K. Aerosol enrichment of the surfactant PFO and mediation of the water-air transport of gaseous PFOA. *Environ. Sci. Technol.* **2008**, *42*, 3969-3974.
- [25] Ochoa-Herrera, V.; Sierra-Alvarez, R. Removal of perfluorinated surfactants by sorption onto granular activated carbon, zeolite and sludge. *Chemosphere* **2008**, *72*, 1588-1593.
- [26] Rayne, S.; Forest, K. An assessment of organic solvent based equilibrium partitioning methods for predicting the bioconcentration behavior of perfluorinated sulfonic acids, carboxylic

acids, and sulfonamides. Nat. Prec. **2009**, <http://hdl.handle.net/10101/npre.2009.3256.1>.

[27] Rayne, S.; Forest, K. Congener specific organic carbon normalized soil and sediment-water partitioning coefficients for the C<sub>1</sub> through C<sub>8</sub> perfluorinated alkylsulfonic and alkylcarboxylic acids. Nat. Prec. **2009**, <http://hdl.handle.net/10101/npre.2009.3011.1>.

[28] Rayne, S.; Forest, K.; Friesen, K.J. Estimated congener specific gas phase atmospheric behavior and fractionation of perfluoroalkyl compounds: Rates of reaction with atmospheric oxidants, air-water partitioning, and wet/dry deposition lifetimes. J. Environ. Sci. Health A **2009**, *in press*.

[29] Richey, D.G.; Driscoll, C.T.; Likens, G.E. Soil retention of trifluoroacetate. Environ. Sci. Technol. **1997**, *31*, 1723-1727.

[30] Jing, P.; Rodgers, P.J.; Amemiya, S. High lipophilicity of perfluoroalkyl carboxylate and sulfonate: Implications for their membrane permeability. J. Am. Chem. Soc. **2009**, *131*, 2290-2296.

[31] Hilal, S.; Karickhoff, S.; Carreira, L. Prediction of the solubility, activity coefficient and liquid/liquid partition coefficient of organic compounds. QSAR Comb. Sci. **2004**, *23*, 709-720.

[32] Tetko, I.V. Computing chemistry on the web. Drug Discov. Today **2005**, *10*, 1497-1500.

[33] Tetko, I.V.; Gasteiger, J.; Todeschini, R.; Mauri, A.; Livingstone, D.; Ertl, P.; Palyulin, V.A.; Radchenko, E.V.; Zefirov, N.S.; Makarenko, A.S.; Tanchuk, V.Y.; Prokopenko, V.V. Virtual computational chemistry laboratory: Design and description. J. Comput. Aided Mol. Des. **2005**, *19*, 453-463.

[34] Tetko, I.V.; Tanchuk, V.Y. Application of associative neural networks for prediction of lipophilicity in ALOGPS 2.1 program. J. Chem. Inf. Comput. Sci. **2002**, *42*, 1136-1145.



- [35] Tetko, I.V.; Tanchuk, V.Y.; Villa, A.E. Prediction of n-octanol/water partition coefficients from PHYSPROP database using artificial neural networks and E-state indices. *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 1407-1421.
- [36] Tetko, I.V.; Poda, G.I. Application of ALOGPS 2.1 to predict log D distribution coefficient for Pfizer proprietary compounds. *J. Med. Chem.* **2004**, *47*, 5601-5604.
- [37] Tetko, I.V.; Bruneau, P. Application of ALOGPS to predict 1-octanol/water distribution coefficients, logP, and logD, of AstraZeneca in-house database. *J. Pharm. Sci.* **2004**, *93*, 3103-3110.
- [38] Weininger, D. SMILES: A chemical language and information system. 1. Introduction to methodology and encoding rules. *J. Chem. Inf. Comp. Sci.* **1988**, *28*, 31-36.
- [39] Weininger, D.; Weininger, A.; Weininger, J.L. SMILES. 2. Algorithm for generation of unique SMILES notation. *J. Chem. Inf. Comp. Sci.* **1989**, *29*, 97-101.
- [40] Rayne, S.; Forest, K.; Friesen, K.J. Congener-specific numbering systems for the environmentally relevant C<sub>4</sub> through C<sub>8</sub> perfluorinated homologue groups of alkyl sulfonates, carboxylates, telomer alcohols, olefins, and acids, and their derivatives. *J. Environ. Sci. Health A* **2008**, *43*, 1391-1401.
- [41] Burns, D.C.; Ellis, D.A.; Li, H.; McMurdo, C.J.; Webster, E. Experimental pK<sub>a</sub> determination for perfluorooctanoic acid (PFOA) and the potential impact of pK<sub>a</sub> concentration dependence on laboratory-measured partitioning phenomena and environmental modeling. *Environ. Sci. Technol.* **2008**, *42*, 9283-9288.
- [42] Rayne, S.; Forest, K.; Friesen, K.J. Computational approaches may underestimate pK<sub>a</sub> values of longer-chain perfluorinated carboxylic acids: Implications for assessing environmental

and biological effects. J. Environ. Sci. Health A **2009**, *44*, 317-326.

[43] Rayne, S.; Forest, K.; Friesen, K. Extending the semi-empirical PM6 method for carbon oxyacid pK<sub>a</sub> prediction to sulfonic acids: Application towards congener-specific estimates for the environmentally and toxicologically relevant C<sub>1</sub> through C<sub>8</sub> perfluoroalkyl derivatives. Nat. Prec. **2009**, <http://hdl.handle.net/10101/npre.2009.2922.1>.

[44] Rayne, S.; Forest, K. ADME/Tox WEB in silico predictions of longer chain perfluoroalkyl carboxylic acid pK<sub>a</sub> values are more accurate than other computational methods. Nat. Prec. **2009**, <http://hdl.handle.net/10101/npre.2009.2936.1>.

[45] Kah, M.; Brown, C.D. Log D: Lipophilicity for ionisable compounds. Chemosphere **2008**, *72*, 1401-1408.

[46] Kelly, B.C.; Ikonomou, M.G.; Blair, J.D.; Surridge, B.; Hoover, D.; Grace, R.; Gobas, F.A.P.C. Perfluoroalkyl contaminants in an arctic marine food web: Trophic magnification and wildlife exposure. Environ. Sci. Technol. **2009**, *in press*, DOI:10.1021/es9003894.

[47] Schwarzenbach, R.P.; Gschwend, P.M.; Imboden, D.M. *Environmental organic chemistry*; John Wiley and Sons: New York, NY, USA, 2003.

[48] Rayne, S.; Forest, K.; Friesen, K.J. Examining the PM6 semiempirical method for pK<sub>a</sub> prediction across a wide range of oxyacids. Nat. Prec. **2009**, <http://hdl.handle.net/10101/npre.2009.2981.1>.

[49] Martin, J.W.; Mabury, S.A.; Solomon, K.R.; Muir, D.C.G. Dietary accumulation of perfluorinated acids in juvenile rainbow trout (*Oncorhynchus mykiss*). Environ. Toxicol. Chem. **2003**, *22*, 189-195.

[50] Martin, J.W.; Mabury, S.A.; Solomon, K.R.; Muir, D.C.G. Bioconcentration and tissue

distribution of perfluorinated acids in rainbow trout (*Oncorhynchus mykiss*). Environ. Toxicol. Chem. **2003**, 22, 196-204.

[51] Rayne, S.; Forest, K.; Friesen, K.J. Estimated bioconcentration factors (BCFs) for the C<sub>4</sub> through C<sub>8</sub> perfluorinated alkylsulfonic acid (PFSA) and alkylcarboxylic acid (PFCA) congeners. J. Environ. Sci. Health A **2009**, 44, 598-604.