Screening Tools for the Bioconcentration Potential of monovalent organic ions in fish

Kai Bittermann[a,c], Lukas Linden[a,c], Kai-Uwe Goss[a,b,*]

[a] Helmholtz Centre for Environmental Research UFZ, Department of Analytical Environmental Chemistry, Permoserstr. 15, D-04318 Leipzig, Germany

[b] University of Halle-Wittenberg, Institute of Chemistry, Kurt-Mothes-Str. 2, D-06120 Halle, Germany

[c] shared first authorship

* corresponding author

Abstract

Currently the bioaccumulation potential of organic chemicals is assessed in a first tier approach via their octanol-water partition coefficient.. This approach has been developed for neutral chemicals and cannot work for ionizable and ionic chemicals because the latter have different sorption-mechanisms and -preferences. Thus, suitable screening tools for the bioconcentration potential of ionic and ionizable chemicals need to be developed because it cannot be expected that these chemicals are non-bioaccumulative per se. Here, we present such screening tools for monovalent ions and ionizable chemicals based on calibrated sorption models for membrane lipids, structural proteins and albumin. The molecular descriptors used for these models arise from quantum chemical calculations and are based on COSMO-RS theory. When we applied our screening tools to 1839 preselected chemicals from the REACH registration data base, we identified 187 chemicals as potentially bioconcentrating (still ignoring any kind of metabolism). Among these were carbon and sulphur based aromatic and aliphatic acids mostly with a rather high molecular surface area. We hope that this outcome will trigger further research on ion specific sorption mechanisms and lead to a re-evaluation of the bioconcentration potential of ionic chemicals.

1 Introduction

There is no generally accepted approach to estimate the bioaccumulation potential of organic ions [1] – despite the fact that the regulation of organic ions is a prevailing challenge.[2] Ionogenic organic chemicals comprise very diverse structures and chemical classes such as surfactants, pharmaceuticals, some classes of pesticides, poly- or perfluorinated acids[2] as well as ionic liquids.[3] The use of a single and easy to determine threshold value (such as a certain logarithmic octanol-water partition coefficient value), which is applied for neutral chemicals by regulation authorities,[4] will not suffice as a standard criterion to identify the bioaccumulation potential of charged chemicals[1]. Previous work focused on the description of rates of uptake and elimination (including metabolism) to describe the bioaccumulation potential of organic ions, aiming at a holistic picture.[5] While we agree that physiologically based pharmacokinetic modelling is highly needed, we consider the underlying physicochemical parameters, especially the equilibrium sorption coefficients to the different relevant phases, as a major uncertainty in our current knowledge. In previous work, both sorption to proteins as well as sorption to membrane lipids was estimated for organic ions via the respective octanol-water partition coefficient.[5] In our own work we have recently shown that this is not appropriate.[6] In this work we therefore develop mechanistic and semi empirical models to predict such equilibrium sorption coefficients. These can then be used to screen the bioaccumulation potential of organic ions in a first tier approach that still neglects any biotransformation or other kinetics and can thus be seen as a worst case scenario. Here, we use the newly developed predictive tools to provide such a screening of chemicals for their bioconcentration potential in fish and based on a depiction of the major sorption matrices. Analogous to the pharmacokinetic literature,[4,7–9] we assume the following sorption matrices in organisms to be the most relevant for organic ions: membrane lipid, muscle protein (which is our proxy for structural protein), serum albumin (which is our proxy for plasma proteins in fish) and water. We have to note that not all fish species have albumin and also there are other blood constituents that might be important sorbents for organic ions. For organic ions we assume that the sorption capacity of storage lipid (fat) can be neglected, based on the finding that ions partition into octanol only marginally as ion pairs[10,11] and octanol is a pretty good proxy for storage lipid,[4] within our general model uncertainties.

While our general approach is straight forward and not new[4] (and has been applied by us before for a few selected ionizable chemicals with available experimental data),[12] the challenge lies in providing all the different partition coefficients required for a broad screening. For the neutral species of ionizable chemicals the usage of poly parameter free energy relationships (pp-LFERs) is an appropriate way to obtain these data as shown in a recent review.[13] In general, pp-LFERs are capable of describing the equilibrium partitioning of neutral organic chemicals between a multitude of biologically relevant matrices and water as well as technical partitioning systems and water. Unfortunately, the applicability of pp-LFERs for ionic organic chemicals is still in its infancy and of rather empirical nature, limited to few chemical classes.[6,14,15] Thus, we investigate here, to what extend required partition coefficients can be estimated with the help of the commercial software COSMOthermX17[1], which is the only predictive tool that cannot only handle neutral species but that is principally able to provide meaningful predictions for the partitioning of organic ions.[6] COSMOtherm is based on quantum mechanical (QM) calculations and fundamental fluid phase thermodynamics (namely the conductor like screening model for real solvents, COSMO-RS)[16,17] which operates with only very general fitting parameters. The COSMO-RS implementation within COSMOtherm is principally applicable to both neutral chemicals as well as ions.[18] For ions, it has particularly been shown to be a good model for the description of the membrane-water partition coefficient [19] and for ionic liquid properties.[20]

Out of the four sorption matrices, only the membrane and water itself are well-defined and are thus directly describable within COSMOtherm.[19,21] The other two important sorption matrices are structural proteins and plasma proteins. About 10% of the whole body mass of vertebrates is made of structural proteins, which themselves consist to about 50% of muscle proteins (e.g., actin and myosin), while the other half is mostly keratin and collagen.[14] In the case of blood plasma the composition of the sorbing matrix varies in different organisms and the contributions of specific proteins are not always clear. Here, we used albumin, which is expected to dominate anionic sorption in human blood, as a proxy for the plasma proteins. For structural proteins and albumin, the only chance to grasp the major characteristics of the respective sorption matrices with COSMOtherm is via fitting experimental partition coefficients of organic ions to so-called sigma moments via a multiple linear regression (MLR). The sigma moments are an output of the quantum chemical cosmo calculation for molecules and account for the solutes' interaction

1 1 Eckert F, Klamt A. COSMOtherm.COSMOlogic, Leverkusen, Germany

properties. Calibrating MLR models based on sigma moments with experimental equilibrium partitioning data works well for neutral chemicals as has been shown for a big variety of liquid-liquid partitioning systems (personal communication COSMOlogic) and is conducted in exact analogy to the pp-LFER approach, as outlined in detail below. We tested this approach both for the partitioning of organic ions between plasma protein and water and structural protein and water.

The major aims of this work were twofold: to develop reliable predictive sorption models, for neutral and ionic chemicals in order to describe the bioaccumulation potential of organic ions and ionizable chemicals (without metabolism); second to identify potentially bioaccumulative compoounds by applying our models to a set of almost 2000 organic ions or ionizable chemicals. For the first aim we developed MLRs based on sigma moments describing the sorption to structural proteins and to albumin (for neutral and monovalent ionic chemicals, respectively). For our second aim, we combined these MLRs with the pp-LFER models for neutral species and COSMOmic for neutral and ionic chemicals and applied it to almost 2000 chemicals.

Materials and Methods for the development of sorption model

Materials for the development of sorption model

Temperature dependence of sorption coefficients

The experimental sorption data for phospholipid membrane are available for temperatures between 20 to 37°C. The sorption differences within this temperature range are negligible, as long as the membrane is in its natural liquid crystalline state.[22] The data for structural and muscle proteins and albumin had been measured at 37°C because they originally aimed to describe sorption capacities in humans. Although the modeled fish has a temperature between 13 and 17°C,[23] we expect only little influence of the temperature dependence of the sorption coefficients and regard this as one of the minor uncertainties of our model.

Sorption to structural (muscle) proteins

Structural proteins such as muscle protein is abundant in vertebrates and of polar nature.[14] Analogous to previous work,[14,24] we assumed the experimental sorption data from water to chicken muscle to be a generally valid proxy for the partition coefficient between structural proteins and water, $K_{\text{structural proteins/water}}$, for both the ionic as well as the neutral species. In fact, for

40 neutral chemicals it has been shown in previous work that the differences in $K_{\text{structural proteins/water}}$ between chicken, fish and pig muscle proteins were small.[24] We used the experimental partitioning data from,[14,24] comprising 63 neutral chemicals, 41 anions and 10 cations (we left out those values that are only given as lower border). In order to be used in our screening model the experimental values had to be converted into volume based partitioning coefficients (multiplied with the density of muscle protein of 1.36 kg/L).[25] Note that this is a rather limited dataset of chemicals. Increasing predictive errors have to be expected for chemicals that do not fall into the range spanned by the calibration data. Given that there are only ten cations in the dataset, a meaningful MLR for cations is not possible (i.e., overfitting is inevitable). This gap needs to be filled by future work. For the time being it might be advisable for the screening to just use a log $K_{\text{structural proteins}}$ (cation) value of 1.5 for any cationic chemical (being the mean value of the exisiting experimental data).

Sorption to albumin

The partitioning to blood plasma is dominated by the sorption to the plasma proteins. Among these proteins serum albumin is the major sorption matrix for both neutral and ionic chemicals.[26,27] We rely on two consistent experimental datasets[26,27] for our model development. The experimental data were derived with bovine serum albumin, which is comparable to human serum albumin.[27] Due to the lack of reliable partitioning data for rainbow trout albumin, we use bovine serum albumin as a surrogate. Obviously, this assumption needs to be revised when new experimental values for fish plasma protein come up and as the circumstances require, a new MLR will have to be set up. In order to be used in our screening, the experimental values were converted to volume based partition coefficients (i.e., they were multiplied with the density of serum albumin, being 1.36 kg/L).[25]

Methods for the development of the sorption model

Calculation of log $K_{\text{fish/water}}$

The partitioning of a permanently charged ionic chemical between any organism and water can be described as the additive sorption to all the sorption matrices in the body of the organism. For ions this is expressed in the following equation for the partitioning into fish:

(1)

with f_x denoting the volume fractions of the respective matrices/phases and the K's describing the

partition coefficients between the matrices/phases and water given in the subscripts (trivially, $K_{water/water}$ equals one and thus only f_{water} needs to be considered). For our screening approach we looked at a 1 kg rainbow trout with the following composition (volume %): storage lipid 11%, phospholipids 1.0 %, structural proteins 15.8 %, plasma proteins 0.27 %, and water 69.8 % (adapted from Nichols et al.).[28] A side note to the wording used here: a 'phase' is per definition homogeneous like water or hexadecane. Phospholipids and albumin are highly heterogeneous, while muscle protein is probably a little less heterogeneous[14] – therefore we denote these latter sorption media as (sorption) matrices.

When we describe the bioaccumulation potential of acids and bases that are partly neutral at the investigated pH, then the partitioning of both species needs to be assessed. For the neutral species, we also consider storage lipids (triglycerides) as a major sorbing compartment in addition to membranes, structural proteins and albumin.[4]

(2)

The total partition coefficients of both species are then combined according to their fractionation in water that depends on the respective pK_a value.

(3)

Note again, that this model is purely based on equilibrium partitioning and does not account for any kind of metabolism and kinetics.

Predicting $K_{x/water}$ for neutral chemicals with pp-LFERs

The partitioning of neutral chemicals to the different sorption phases/matrices listed in Eq. 2 can be predicted with poly parameter free energy relationships (pp-LFERs) from the literature. In general, pp-LFER models are widely used and accepted as documented by a number of reviews.[13,29,30] We used the UFZ-LSER database[31] in order to get a maximum amount of experimentally determined solute descriptors, L (log of the hexadecane-air partition coefficient), S (dipolarity/polarizability parameter), A (solute H-bond acidity), B (solute H-bond basicity), and V (molar volume). For cases where no experimental solute descriptors were available we used the UFZ-QSPR, available free of charge from the same source. We used these solute descriptors in the following pp-LFERs from the literature to calculate $K_{membrane/water}$(neutral), $K_{storage\ lipid/water}$(neutral), $K_{structural\ proteins/water}$(neutral), and $K_{albumin/water}$(neutral) respectively[22,24,26,32]:

; n=131, SE=0.28, T=37°C (4)

; n=247, SE=0.20, T=37°C (5)

; n=46, SE=0.23, T=37°C (6)

; n=82, SE=0.41, T=37°C (7)

In addition to the pp-LFER predictions, the partitioning of neutral chemicals to structural proteins and albumin were also predicted with multi-linear regressions (MLRs) against the sigma moments of the respective chemicals (as outlined in detail below), while the partitioning of neutral chemicals to membrane was also predicted with COSMOmic. Hence, for the neutral chemicals we ended up having two predictive models (based on the same calibration data sets) one using the pp-LFER approach and one using the sigma moments derived from quantum chemical cosmo calculations (see below). We expect that all models have their shortcomings due to the finite training set, so we decided to use a consensus model for neutral chemicals, meaning that $K_{x/water}$ of the respective sorption matrix was finally determined by the average of the two respective model results. For storage lipid we relied solely on the ppLFER Eq. 5.

Generation of COSMOfiles

Prior to the partitioning calculations with COSMOtherm (including the calculations via COSMOmic or via sigma moments) COSMOfiles of the respective chemicals were generated with quantum mechanical calculations (BP-TZVP level):[33–35] We used COSMOconfX16 and Turbomole version 7.1 for full energy minimization and conformer generation (up to ten conformers were generated).[36]

Predicting $K_{x/water}$ of ionic and neutral chemicals via sigma moments

Analogous to the pp-LFER approach which uses 5 solute descriptors (called Abraham descriptors) the interaction possibilities of a solute can also be described with five descriptors, derived from the COSMOfile of the specific chemical. In fact, it has been demonstrated that the five Abraham solute descriptors for neutral chemicals correlate well with the following five sigma moments Sig0, Sig2, Sig3, Hb_acc3 and Hb_don3 – all of which can be calculated with the commercial software COSMOtherm.[37] Given that a) these five sigma moments are also well-suited for describing partitioning for neutral chemicals via a multi-linear regression (MLR) [37,38] and b) the partitioning systems of structural protein and plasma protein cannot directly be

modelled with COSMOtherm it is an obvious choice to use the sigma moments to describe the respective partitioning systems with a MLR of the following general form:

(8)

This is done in exact analogy to the calibration of a pp-LFER equation – but unlike the pp-LFERs, sigma moments should per se be able to describe both ionic and neutral chemicals, if we also consider the additional sigma moment Sig1, which describes the charge. A big advantage of sigma moments based MLR's over other QSAR's is, that the sigma moments describe intuitively understandable physicochemical parameters, as outlined in the SI.

Sorption to structural (muscle) proteins

For $K_{\text{structural proteins/water}}$ (ion) a tentative ppLFER had already been set up for monovalent ions by including additional descriptors accounting for the charge.[14] However, this pp-LFER can only account for the ionic forms of phenols, carboxylic acids, pyridines and amines and is therefore not suited for our screening purpose. Therefore we modeled $K_{\text{structural proteins/water}}$ via the MLR based on sigma moments as discussed above (range of the sigma moments is shown in SI Table 1).

Sorption to albumin

The sorption of ions to serum albumin is partly influenced by strong steric effects,[14] which can only be included in a modelling approach through extensive calibration and calculation effort.[25] Such a model is not feasible for our screening purpose because a) it requires a very time-consuming and meticulous calculation effort and b) its domain of applicability is rather narrow.[25] But we can use the existing experimental data for a simplified model (which is expected to have a wider applicability domain while predicting the fitting data set less accurate) that is based on the sigma moments as discussed above. Prior to construction of this sigma-moment based model we excluded those chemicals whose sorption behaviour to albumin is highly influenced by steric effects, which cannot be covered by the sigma moments. Due to our previous 3D-QSAR modeling experience[25] we know that especially anions that have a substitution in direct vicinity to the carboxylic group are strongly influenced by steric effects (they experience a twist of the carboxyl group). Thus, we excluded these anions from the calibration dataset,[27] namely 2,6-dichlorobenzoic acid anion, 2-chlorobenzoic acid anion, 2-naphthalenacetic acid anion, 2-naphthoic acid anion, and naphthalene-2-sulphonate anion. We have to note here that there are

not enough experimental data to provide reliable rules as to which other chemicals would fall outside the application range.

Sorption to membrane lipid

$K_{membrane/water}$(ion) of ionic organic chemicals can be modeled with the COSMOmic application in COSMOtherm,[19] which is currently the most reliable method available for this purpose[6] and the only prediction method that can be used for screening purposes[39] (in contrary to MD simulations). COSMOmic has been validated with a rather diverse dataset, including a few zwitterions and di-cations. For our screening approach, we used exactly the same calculation details as in the original COSMOmic publication: 1401 parametrization of the COSMOtherm software with an offset of 0.32 log units for the prediction of $K_{membrane/water}$ of organic ions, using a pure DMPC membrane.[19]

For neutral chemicals the sorption to membrane lipid, $K_{membrane/water}$(neutral) was also modeled with COSMOmic (with the same settings as used for ions), and, additionally, with the ppLFER shown in Eq. 4.

Overall Workflow

Once, all predictive models for the required partition coefficients had been set up, we were able to start the screening task. Our overall Screening workflow can be summarzized as outlined in Fig. 1.

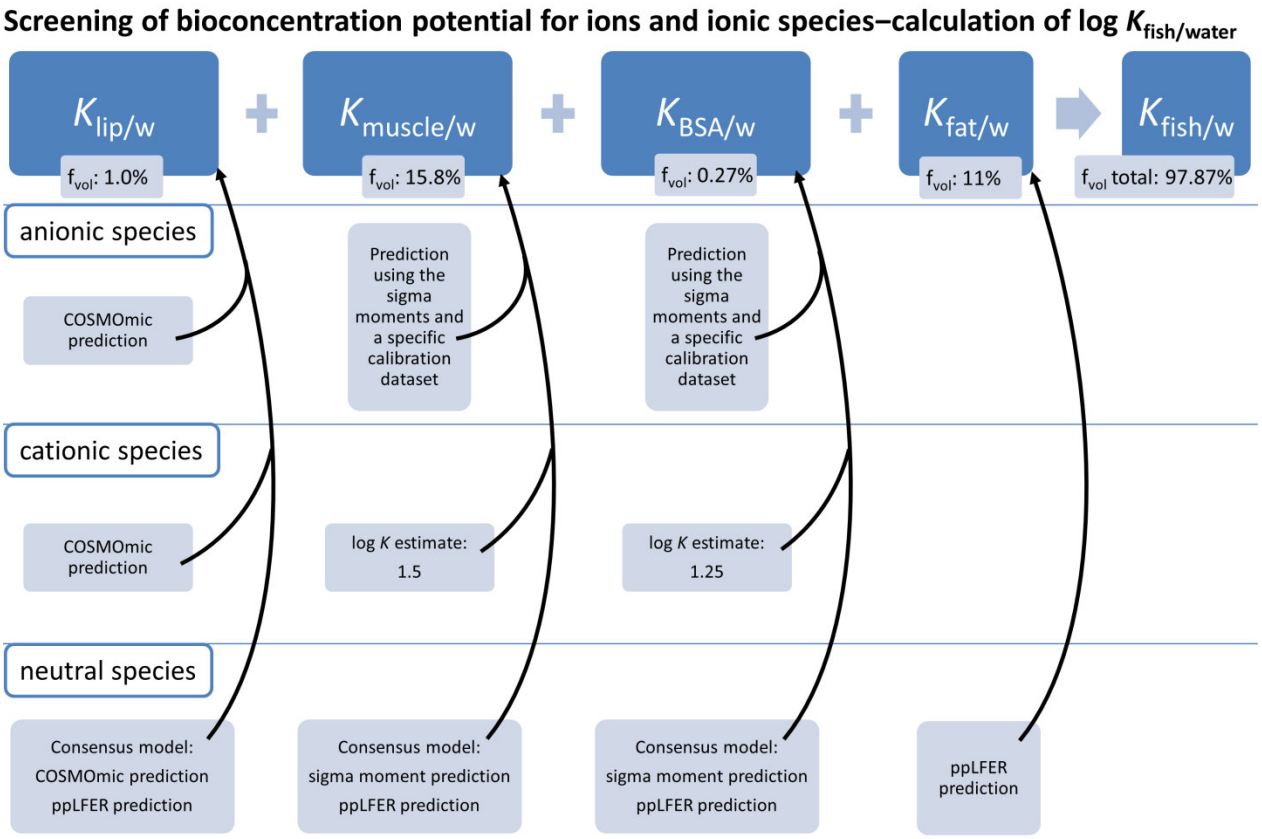


Figure 1: Workflow for our screening procedure for potentially bioaccumulative chemicals. All models based on MLRs with sigma moments were newly developed in this work.

Screening for potentially bioaccumulative chemicals

For our screening for potentially bioaccumulative ions we largely investigated a dataset provided by ECHA of more than 70000 non-confidential chemical structures from the REACH registration database and Classification and Labelling inventory. We first filtered the dataset for those chemicals with a molecular weight between 100 and 800, having only one pK_a. For chemicals with a pK_a between 3 and 7 both, the neutral and anionic species were considered. Here, we relied on the pK_a's given in the ECHA dataset, which were predicted with the ChemAxon software package. If the pK_a's were below 3, we only considered the anionic fraction, if the pK_a's were above 11 we only considered the cationic species. Also, we restricted our investigation on chemicals constructed by the atoms H, C, N, S, O, P and halogenates. We further included some chemicals in our screening of known environmental relevance such as perfluorinated chemicals, ionic liquids and quaternary phosphonium cations. If adequate, we predicted the pK_a of these chemicals with JChem for Excel, version 15.10.2600.341 (Copyright 2008-2015 ChemAxon Ltd. https://www.chemaxon.com/)

using a SMILES code as input. According to the literature, JChem performs equally well as ACD and the topological method MoKa on pKa predictions.[40]

Results and Discussion

Models for the different sorption matrices

Structural protein

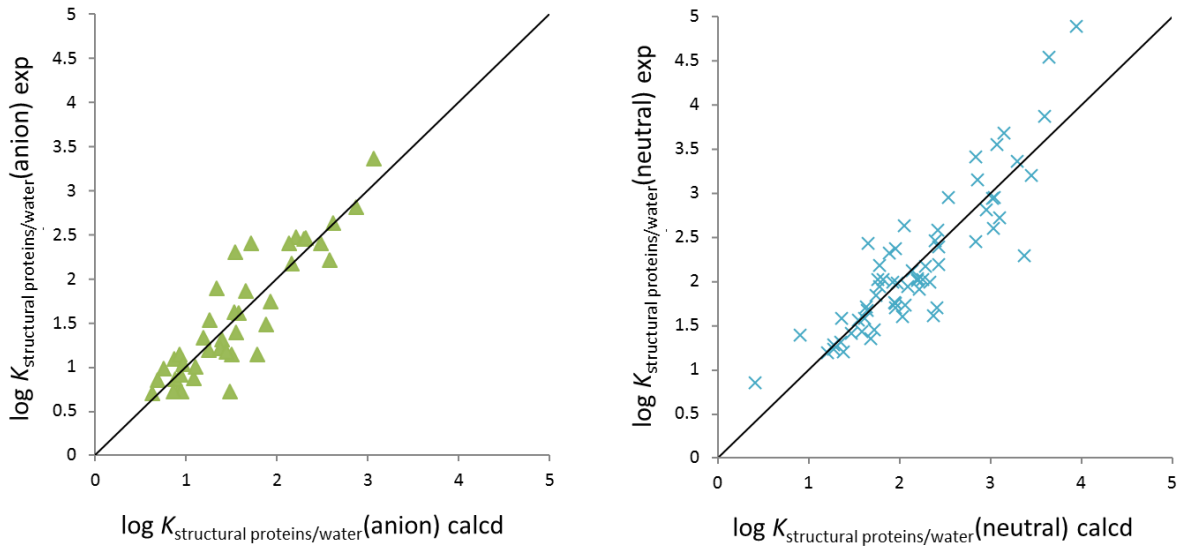
The Abraham solute descriptors are always positive (with the notable exception of perfluorinated chemicals and silicates), which makes the resulting pp-LFER equation instructive and easily understandable.[13] In contrast, Sig 1 and Sig3 can also take on negative values. This and the fact that the absolute values of the sigma moments are not normalized prohibit an easy interpretation of MLRs fitted with sigma moments as compared to pp-LFER equations.

In a first attempt, we fitted the experimental data of the 63 neutral chemicals, 41 anions and ten cations altogether with a MLR and obtained a promising correlation already (RMSE = 0.46, R^2 = 0.67, SI Fig. 2). But we also assumed that differently charged chemicals might sorb to different sorption sites within the muscle proteins, so we also fitted the neutral chemicals and anions with MLRs separately. These two fits have less fitting parameters because we excluded those parameters that had a standard deviation larger than the fitted parameters themselves (resulting in three sigma moments and one constant for the anions, and four sigma moments and one constant for neutral chemicals). Additionally, the separate fits had a better statistical outcome (i.e., the RMSE was smaller and R^2 was higher) and are thus our first choice for screening.

; R^2= 0.81, RMSE=0.30, F=53, n=41 anions (9)

,

R^2= 0.78, RMSE=0.38, F=52, n=63 neutral chemicals (10)



289

290 **Figure 2: MLR based on sigma moments for structural protein (chicken muscle), left for**
291 **anions (3 descriptors + constant),** RMSE = 0.30, R^2 = 0.81, **right for neutral chemicals (4**
292 **descriptors + constant),** RMSE = 0.38, R^2 = 0.78.

293

294 Unfortunately, for cationic chemicals, there were only ten data points; we regard this as not
295 enough for a meaningful MLR. Therefore we decided to add the average
296 log $K_{structural\ proteins/water}$(cation) value of 1.5 for cations for screening purposes as a rough estimate
297 (originating from the ten cations of the dataset and their log $K_{albumin/water}$(cation) range of 0.97 to
298 2.29).

299 For neutral chemicals the prediction of $K_{structural\ proteins/water}$(neutral) is also possible with a pp-LFER
300 equation.[24] Analogous to the calculation of $K_{membrane/water}$(neutral) we used a consensus model for
301 the neutral chemicals, averaging the outcomes of Eq.s 6 and 10.

302 **Albumin**

303 Analogous to the structural protein, it is plausible to assume that anions and neutral chemicals
304 sorb to different sorption sites within the BSA protein. This can explain the rather poor fit of the
305 data, when the 40 anions and the 83 neutral chemicals are fitted together (SI Fig 3). The separated
306 fits of anions and neutral chemicals yield the following system descriptors (again leaving out
307 insignificant descriptors).

308 ; R^2= 0.82, RMSE= 0.33, F= 39, n=40 anions (11)

Note that this model has to be used with caution for those anionic chemicals that are sterically hindered in vicinity to a carboxyl group, as explained above.

; R^2= 0.56, RMSE= 0.57, F= 25, n=83 neutral chemicals (12)

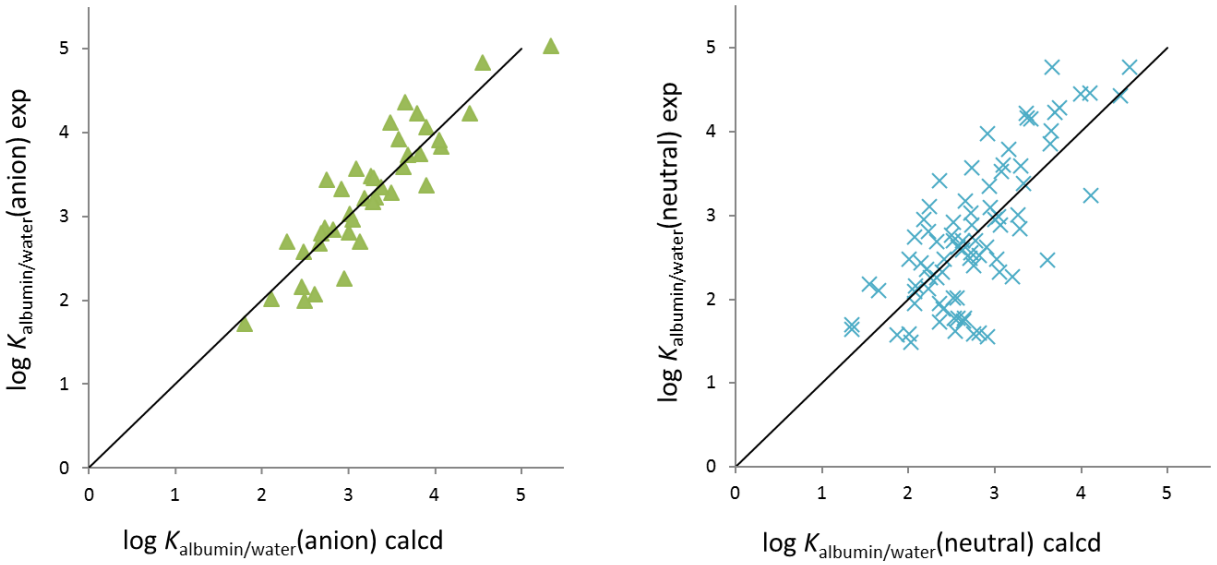


Figure 3: MLR based on sigma moments for albumin, left for anions (4 descriptors + constant), right for neutral chemicals (4 descriptors + constant).

Again, we described the partitioning to BSA for neutral chemicals with a consensus model, averaging the results from Eq. 12 and the ppLFER Eq. 7 (SI Fig. 4).

As before, there are not enough data for cations to establish a MLR, so we used the average log K value of 1.25 (originating from the four cations of the dataset and their log $K_{albumin/water}$(cation) range of 0.97 to 1.58).

Model constraints

In order to facilitate the interpretation of the results and prevent misuse of the model, we repeat the model weaknesses in a bullet point form here.

- It is questionable whether poly- and perfluorinated chemicals are well-described with the sigma moment approach, given that van der Waals interactions are only depicted via the Sig1 (area). We therefore expect systematic deviations for perfluorinated chemicals, but due to the lack of experimental data this cannot be quantified.

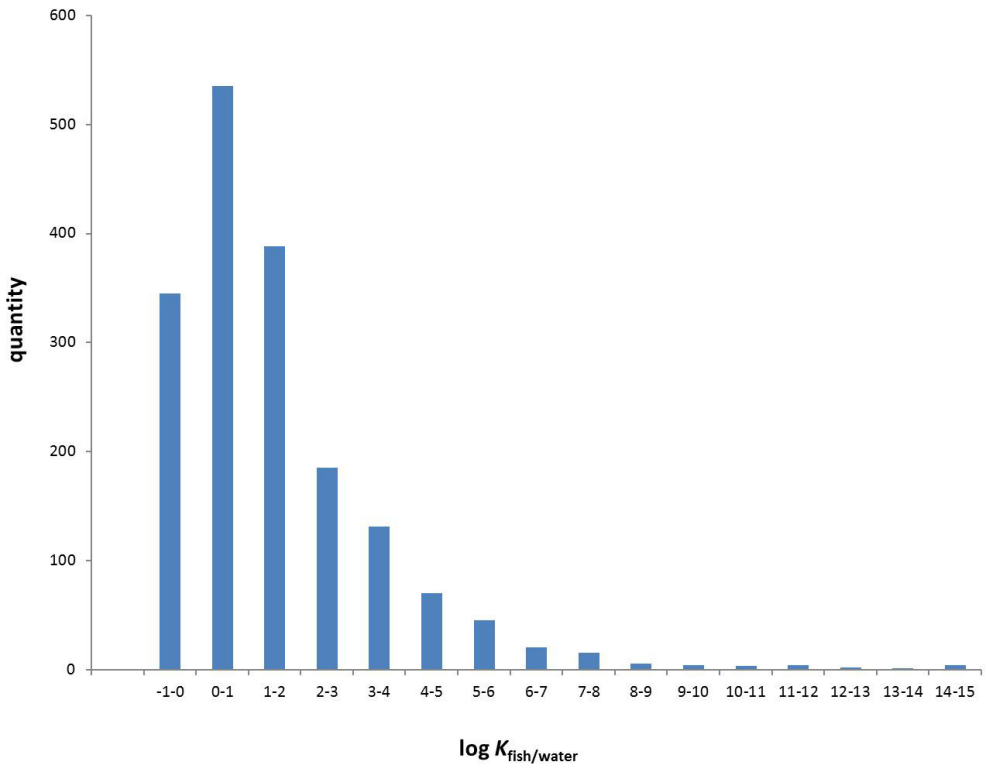
- Unfortunately, also for neutral chemicals there is a lack of experimental data for perfluorinated chemicals. So also the pp-LFER based submodels for neutral chemicals can only be used with great caution for this class of chemicals

- Sorption of cations to structural proteins and plasma proteins is only roughly estimated by average values due to an insufficient number of calibration data (i.e., the sorption to serum albumin is presumably weak for cations but they sorb stronger to other plasma proteins than albumin which are not included in our screening approach due to the lack of consistent data).[41]

- Complex ions, i.e., ions with several ionizable groups, as well as surfactants were not part of the calibration or validation set of our models and the model performance for these chemicals/species is unknown.

- Chemicals that show a distinct steric effect in their sorption to serum albumin might not be correctly covered by our modelling approach.

Screening of potentially bioaccumulative monovalent organic ions



342

343 **Figure 4: Histogram of calculated log $K_{fish/water}$ according to Eq. 3.**

344 We screened 1839 preselected chemicals for their bioaccumulative potential, 187 (10%) of them
345 have predicted log $K_{fish/water}$ values larger than 4 (Fig 4, see SI Table 3 for the dominating sorption
346 phase). The molecular weight of these potentially bioaccumulative chemicals ranged between
347 255 and 756 *u* thus spanning almost the entire range of the preselected values (see methods).

348

349 **Figure 5: Calculated log $K_{fish/water}$** (combining the contribution from the neutral and ionic species)
350 against Sig0 (area) .

351

352 Log $K_{fish/water}$ correlates reasonably with the molecular surface area as it can be expected (Fig 5);
353 larger chemicals tend to be more bioaccumulative than smaller ones due to their increased
354 hydrophobicity. For the chemicals that possess a neutral and anionic species at a pH of 7 (acids),
355 the neutral species has generally the higher log $K_{fish/water}$ value compared to the anionic species (SI
356 Fig. 7). But, we also compared the pH dependent contribution of the two species to $K_{fish/water}$(total)

and in most of the cases the anionic species dominated the $K_{fish/water}$(total) at pH 7 (SI Fig. 8). The two outliers in Fig. 5 with a relatively high log $K_{fish/water}$ but a rather small Sig0 of roughly 200 are adamantanes, which are cubic molecules with a relatively small volume.

For the further discussion, we only consider the contribution and the influence of the ionic species on the bioaccumulative potential because this is the most important contribution.

Table 1: Overview of screened chemicals. Note that the sum of the chemicals in the sub-groups does not always add up to the total of 1839 chemicals, because not always all chemicals fit into the shown categories.

			log $K_{fish/water}$ < 4		log $K_{fish/water}$ > 4	
			quantity	%	quantity	%
total		**1839**	**1652**	**89.83%**	**187**	**10.17%**
aromatic		822	783	95.26%	39	4.74%
aliphatic		942	794	84.29%	148	15.71%
S based acid		409	343	83.86%	66	16.14%
C based acid		606	562	92.74%	44	7.26%
Sorbing matrix dominated by structural proteins		177	165	93.22%	12	6.78%
dominated by plasma proteins		408	354	86.76%	54	13.24%
	S based acid	102	75	73.53%	27	26.47%
	C based acid	223	207	92.83%	16	7.17%
	aromatic	250	228	91.20%	22	8.80%
	aliphatic	158	207	131.01%	32	20.25%
dominated by membrane lipids		266	171	64.29%	95	35.71%
	S based acid	34	12	35.29%	22	64.71%
	C based acid	33	17	51.52%	16	48.48%
	aromatic	66	59	89.39%	7	10.61%
	aliphatic	200	112	56.00%	88	44.00%

Figure 6 log $K_{Fish/water}$ ionic against the surface area of the ionic chemical. The color code indicates the dominating sorption matrix that contributes for more than 60% of the total log K value.

Analysis of the results indicated the following general trends :

- aliphatic chemicals tend to be more bioaccumulative than aromatic chemicals (Tab. 1)

- Sorption to albumin is generally dominated by smaller chemicals while bigger molecules tend to sorb stronger to membrane lipids

- Structural proteins as a dominating sorption matrix (>60%) rarely leads to bioconcentration potential (Fig. 6)

- Albumin and membranes dominate the sorption behaviour of bioconcentrating chemicals, (Fig. 6).

- aromatic chemicals preferably sorb to albumin while aliphatic chemicals preferably sorb to membranes (Table 1)

- sorption of S based acids is dominated by plasma proteins while for C based acids sorption can be dominated by both plasma proteins and membranes (Table SI 4).

- Sorption to plasma proteins is similar to sorption to structural proteins but higher (SI Fig. 9)

- Sorption to plasma proteins is considerably different to sorption to membranes but high values correlate (SI Fig. 10)

Implications for the regulatory process

Assessment of the bioconcentration potential of ionizable organic chemicals often assumes that ionic species do not partition into biological matrices. However, there are sufficient data in the literature to show that this general assumption does not hold. Hence, it is inacceptable to waive the bioconcentration potential of organic ions based on this assumption.. Here, we have shown that a first tier screening of ions is possible, based on molecular descriptors that came from quantum chemical calculation. It must be noted that our assessment was based on the assumption that the exposure pH value is the same as the internal pH of the fish (i.e., pH=7). If this is not the case then an ion-trap effect will occur which further increases the bioconcentration potential for organic acids [42]. If the pH in the exposure medium is 2 log units smaller (i.e., pH= 5) then the BCF increases by a factor 100 for acids with a pKa < 5 due to ion trapping.

We also point out that the presented screening was only directed towards bioaccumulation in fish. Many of the chemicals that are not expected to have a BCF potential may still have a substantial

bioaccumulation potential in terrestrial organisms. Chemicals that are not volatile (log K_{oa} > 5) and that do not metabolize, possess a bioaccumulative potential in terrestrial organisms if their log $K_{organism/water}$ exceeds 1 [43–46]. If we take the $K_{fish/water}$ as a proxy for a more general $K_{organism/water}$ for all vertebrates, then a large portion of the ionizable chemicals tested here would be classified as potentially bioaccumulative in air breathing organisms. Of course, any initial screening for partition properties has to be followed by more specific testing that also considers biotransformation in the test organisms.

Conflicts of interest

There are no conflicts to declare

Acknowledgement

We thank Jane Caley from ECHA for providing us with **the** data set on ionic chemicals from the registration database. We thank Andreas Klamt and his Co-workers of COSMOlogic for supporting this work with their COSMO-database and their calculations of new COSMO files on the BASF supercomputer. We also thank the BASF for providing computation time free of charge. Financial support from the German Environment Agency under project FKZ 3717 67 402 0 is also acknowledged.

References

1 G. Treu, W. Drost, U. Jöhncke, C. Rauert and C. Schlechtriem, *Environ. Sci. Eur.*, 2015, **27**, 34.

2 A. Franco, A. Ferranti, C. Davidsen and S. Trapp, *Int. J. Life Cycle Assess.*, 2010, **15**, 321–325.

3 T. P. Thuy Pham, C.-W. Cho and Y.-S. Yun, *Water Res.*, 2010, **44**, 352–372.

4 S. Endo, T. N. Brown and K.-U. Goss, *Environ. Sci. Technol.*, 2013, **47**, 6630–9.

5 J. M. Armitage, J. A. Arnot, F. Wania and D. Mackay, *Environ. Toxicol. Chem.*, 2013, **32**, 115–128.

6 K. Bittermann, S. Spycher and K.-U. Goss, *Chemosphere*, 2016, **144**, 382–391.

7 P. Poulin and F. P. Theil, *J. Pharm. Sci.*, 2000, **89**, 16–35.

8 W. Schmitt, *Toxicol. In Vitro*, 2008, **22**, 457–467.

9 T. Rodgers and M. Rowland, *J. Pharm. Sci.*, 2006, **95**, 1238–1257.

10 B. I. Escher and L. Sigg, in *Physicochemical Kinetics and Transport at Biointerfaces*, eds. H. P. v. Leeuwen and W. Köster, John Wiley & Sons, Ltd, Chichester, UK, 9th edn., 2004, pp. 205–269.

11 B. I. Escher and R. P. Schwarzenbach, *Environ. Sci. Technol.*, 1996, **30**, 260–270.

12 K.-U. Goss, K. Bittermann, L. Henneberger and L. Linden, Chemosphere, 2018, **199,** 174-181..

13 S. Endo and K. Goss, *Environ. Sci. Technol.*, 2014, **48**, 12477–12491.

14 L. Henneberger, K.-U. Goss and S. Endo, *Environ. Sci. Technol.*, 2016, **50**, 7029–7036.

15 M. H. Abraham and W. E. Acree, *J. Chromatogr. A*, 2016, **1430**, 2–14.

16 A. Klamt and G. Schüürmann, *J. Chem. Soc. Perkin Trans. 2*, 1993, 799–805.

17 A. Klamt, *J. Phys. Chem.*, 1995, **99**, 2224–2235.

18 A. Klamt, *Fluid Phase Equilib.*, 2015, **407**, 152–158.

19 K. Bittermann, S. Spycher, S. Endo, L. Pohler, U. Huniar, K.-U. Goss and A. Klamt, *J. Phys. Chem. B*, 2014, **118**, 14833–42.

20 M. Diedenhofen and A. Klamt, *Fluid Phase Equilib.*, 2010, **294**, 31–38.

21 A. Klamt, U. Huniar, S. Spycher and J. Keldenich, *J. Phys. Chem. B*, 2008, **112**, 12148–12157.

22 S. Endo, B. I. Escher and K.-U. Goss, *Environ. Sci. Technol.*, 2011, **45**, 5912–5921.

23 OECD, *Test No. 305: Bioaccumulation in fish and dietary exposure*.OECD publishing, 2012

24 S. Endo, J. Bauerfeind and K. U. Goss, *Environ. Sci. Technol.*, 2012, **46**, 12697–12703.

25 L. Linden, K.-U. Goss and S. Endo, *Environ. Sci. Process. Impacts*, 2017, **19**, 261–269.

26 S. Endo and K.-U. Goss, *Chem. Res. Toxicol.*, 2011, **24**, 2293–2301.

27 L. Henneberger, K. U. Goss and S. Endo, *Environ. Sci. Technol.*, 2016, **50**, 5119–5126.

28 J. W. Nichols, J. M. McKim, M. E. Andersen, M. L. Gargas, H. J. Clewell and R. J. Erickson, *Toxicol. Appl. Pharmacol.*, 1990, **106**, 433–447.

29 M. H. Abraham, A. Ibrahim and A. M. Zissimos, *J. Chromatogr. A*, 2004, **1037**, 29–47.

30 M. Vitha and P. W. Carr, *J. Chromatogr. A*, 2006, **1126**, 143–194.

31 N. . E. Ulrich S.; Brown, T.N.; Watanabe, N.; Bronner, G.; Abraham, M.H.; Goss, K.-U., 2017.

32 A. Geisler, S. Endo and K. U. Goss, *Environ. Sci. Technol.*, 2012, **46**, 9519–9524.

33 A. D. Becke, *Phys. Rev. A*, 1988, **38**, 3098–3100.

34 J. Perdew, *Phys. Rev. B*, 1986, **33**, 8822–8824.

35 A. Schäfer, C. Huber and R. Ahlrichs, *J. Chem. Phys.*, 1994, **100**, 5829.

36 M. J. Vainio and M. S. Johnson, *J. Chem. Inf. Model.*, 2007, **47**, 2462–2474.

37 A. M. Zissimos, M. H. Abraham, A. Klamt, F. Eckert and J. Wood, *J. Chem. Inf. Comput. Sci.*, 2002, **42**, 1320–1331.

38 C. Mehler, A. Klamt and W. Peukert, *AIChE J.*, 2002, **48**, 1093–1099.

39 K. Bittermann and K.-U. Goss, *Chemosphere*, , DOI:10.1016/j.chemosphere.2017.05.097.

40 J. C. Shelley, D. Calkins and A. P. Sullivan, *J. Chem. Inf. Model.*, 2011, **51**, 102–104.

41 J. M. Kremer, J. Wilting and L. H. Janssen, *Pharmacol. Rev.*, 1988, **40**, 1–47.

42 J. Neuwoehner and B. I. Escher, *Aquat. Toxicol.*, 2011, **101**, 266–275.

43 K.-U. Goss, T. N. Brown and S. Endo, *Environ. Toxicol. Chem.*, 2013, **32**, 1663–71.

44 J. M. Armitage and F. A. P. C. Gobas, *Environ. Sci. Technol.*, 2007, **41**, 4019–4025.

45 B. C. Kelly, M. G. Ikonomou, J. D. Blair, A. E. Morin and F. A. P. C. Gobas, *Science (80-.).*, 2007, **317**, 236–239.

46 G. Czub and M. S. McLachlan, *Environ. Sci. Technol.*, 2004, **38**, 2406–2412.