Chemical similarity: structuring risk and hazard assessment

Wassenaar, P.N.H.

5

Variability in Fish Bioconcentration Factors: Influences of Study Design and Consequences for Regulation

Pim N.H. Wassenaar, Eric M.J. Verbruggen, Ellen Cieraad, Willie J.G.M. Peijnenburg and Martina G. Vijver

Published in Chemosphere 239 (2020), 124731.

Abstract

The fish bioconcentration factor (BCF) is an important aspect within bioaccumulation assessments. Several factors have been suggested to influence BCF values – including species, developmental stage, mixture exposure, and calculation method. However, their exact contribution to variance in BCF values is unknown. Within this study we assessed the relative impact of these test characteristics on BCF values and analyzed the reproducibility of aquatic exposure bioconcentration tests.

Linear mixed effects analyses were performed on a newly develop database to investigate the relationship between the response variable (i.e. lipid normalized log BCF values) and several test characteristics as fixed effects.

Lower BCF values were observed for substances that were simultaneously applied with high molecular weight polycyclic aromatic hydrocarbons compared to single substance exposure (with an average difference of -0.81 log BCF). Also, lower BCFs upon kinetic determination were observed compared to steady-state BCFs (log BCF -0.27), and lower BCFs for species from the Ostariophysi subcohort level (log BCF -0.17 to -0.15). In addition, data analysis showed high variation within BCF values for single substances (average SD = log BCF 0.21), which questions the robustness of the current bioaccumulation assessments. For example, the 95% confidence range of a BCF value of 2500 ranges from 953 ('not-bioaccumulative') to 6561 ('very bioaccumulative').

Our results show that the use of one single BCF leads to a high uncertainty in bioaccumulation assessments. We strongly recommend that within future bioconcentration studies, the used experimental design and test conditions are described in detail and justified to support solid interpretation.

5.1 Introduction

The bioaccumulation potential of chemicals is an important factor within risk assessment. Accumulation may result in high internal concentrations leading to toxicity, even when external concentrations are low [116]. Therefore, substances with a high bioaccumulation potential are of concern, with even higher concerns for substances that – besides being (very) bioaccumulative –are also (very) persistent in the environment and/or toxic to humans or biota (i.e. PBT/vPvB-assessment). From a regulatory point of view, emissions of such substances should be minimized as much as possible, as their effects are unpredictable in the long-term, and as it is very difficult to remove the substances from the environment [19].

International regulatory criteria on bioaccumulation assessment (B-assessment) are mainly based on bioconcentration factors (BCF) in aquatic species [117]. BCFs represent the accumulation of a substance via aquatic exposure, and can be determined under laboratory-controlled conditions via OECD Test Guideline 305, ASTM E1022-94 or OPPTS Test Guideline 850.1730. Within this test, the BCF is determined at steady-state conditions (i.e. the ratio of the substance concentration in fish, C*f*, to the water concentration, C*w*, at steady state) or via kinetic determination (i.e. the ratio of the uptake rate constant, k_1, to the depuration rate constant, k_2). For very hydrophobic substances the BCF could alternatively be determined via dietary exposure [118]. In principle, a substance is considered to be bioaccumulative when the BCF value exceeds a specific threshold. Depending on the regulatory framework, the bioaccumulation cut-off value ranges from 500-5000 (Table 5.1). In addition, within some legislations, bioaccumulation factors (BAFs) or octanol-water partitioning coefficients (log K_{ow}) can also be considered within the B-assessment [117]. The consequences of B-classification varies from product labeling, restrictions in use, to minimization of emissions, with the ultimate aim of chemical substitution (e.g. for PBT/vPvB substances).

Within current regulatory frameworks, one BCF value is generally sufficient to conclude on the bioaccumulative properties. Hence, the variation (i.e. reproducibility) of this value is usually not considered. Several biotic and abiotic factors have been suggested to influence BCF values – including species, developmental stage, exposure method, calculation method, and various others [119]. And although known, the accepted experimental designs often do not specify or take into account such factors, as their exact contributions are unknown. Only some guidance and advice is provided within test guidelines with respect to preferences and/or reporting of these factors [19,118]. However, because of the importance of the B-assessment within chemical safety assessment – as indicated by the relative high number of test requests in Europe [120] – it is considered relevant to analyze the contribution of the factors that are suggested to affect the bioaccumulation potential of chemicals.

Table 5.1: BCF threshold values as applied in several international regulations [117].

Regulation	Assessment type	Bioaccumulative	Very bioaccumulative
POPs UNEP Stockholm Convention	POPs identification	5000	-
OSPAR Convention	PBT substances identification	500	-
Canadian Environmental Protection Act (CEPA)	PBT substances identification	5000	-
US Toxic Substances Control Act (TSCA)	PBT/vPvB substances identification	1000	5000
Australian National Industrial Chemicals Notification and Assessment Scheme (NICNAS)	PBT/vPvB substances identification	2000	5000
EU REACH Regulation (1907/2006)	PBT/vPvB substances identification	2000	5000
EU Plant Protection Product Regulation (1107/2009)	PBT/vPvB substances identification	2000	5000
EU Biocidal Products Regulation (528/2012)	PBT/vPvB substances identification	2000	5000
UN Globally Harmonised System (GHS)	Hazard classification and labelling	500	-
EU CLP Regulation (1272/2008)	Hazard classification and labelling	500	-

On top of that, in recent years, there has been an increasing interest in the development of alternative bioaccumulation tests, as the fish bioaccumulation studies are time consuming, expensive and animal demanding. Several new models include *in silico*, *in vitro* and invertebrate or early-life-stage *in vivo* test systems [121–124]. In order to evaluate their performance, performance information on the reference benchmark, i.e. the aqueous OECD 305 test, is necessary.

In this study we analyze and evaluate the reproducibility and influential factors for the bioconcentration test via aquatic exposure. Using a newly developed database of bioconcentration values, we assessed the impact of different test characteristics (e.g. combination exposure, calculation methodology, species and life stage of the fish) on BCF values and their variation. These test characteristics were selected specifically, because of their potential influence on BCF values and the availability of relevant information in reported studies.

5.2 Methodology

5.2.1 Data selection

Experimental BCF values were selected from the databases as developed by Arnot and Gobas [125] and the Japan METI-NITE database [126] (data extracted on 19-03-2018). Data were restricted to aquatic exposure experiments with fish, only considering direct exposure (i.e. excluding studies investigating bioconcentration in the second generation) and limited to laboratory-derived data. For each experiment only one overall BCF value was included, thus excluding all intermediate measurements. In case of steady-state BCF values, the included value involves the reported BCF value or the average of all BCF values at steady state. In addition, reported BCF values below or above a certain value (i.e. '<' or '>') were excluded as no absolute value was derived. Identified data were scored on reliability based on criteria related to substance concentration, reported BCFs, and general test conditions. The following substance based criteria were used: 1) the water exposure concentration should be measured and not nominal; and 2) water exposure concentrations should be below water solubility limits (as estimated by WSKOW v1.42 from EPISuite [37]). With respect to the reported BCF values, the following criteria were applied: 3) reported BCF values should be substance specific (e.g. not based on total radiolabeled content); 4) when BCF steady-state values are reported, exposure duration needs to be sufficient to reach steady-state conditions (this aspect was analyzed similar as assessed by Arnot and Gobas [125]: when "steady state" was declared by the authors, or when time was sufficient to reach 80% of steady state according to model estimations [125]); 5) the BCF should be based on whole body content; and 6) lipid content of the fish should be reported. In addition, several experimental test conditions should be met. Total organic carbon content must be lower than 2 mg/L, pH should be between 6 and 8.5 at the start of the experiment, temperature must be close to the recommended ranges as reported in the OECD TG 305 [118] and must not be below 3°C or above 30°C, the dissolved oxygen concentration must be above 60% of saturation and no toxicity should be observed during the test. Data was included in case that ranges of organic carbon, pH or oxygen concentrations were reported that partially meet the criteria, or when these parameters were not reported. These quality screening criteria are comparable to those suggested in previous studies [125,127].

For substances with at least one reliable BCF value, we gathered additional data via the OECD QSAR Toolbox [128] and the US-EPA ECOTOX database [129] (data extracted on 02-04-2018). Retrieved BCF values were scored on reliability, similar as described above. Ultimately, only substances with three or more unique BCF values were used for further analysis, and substances with less BCF values were excluded.

5.2.2 Data extraction

For the included data, we collected parameters related to bibliographic data, chemical descriptors, test conditions and endpoint information (see Table S1). *Bibliographic data* includes the first-author, reference and year of publication. Information on the *chemical descriptors* consists of CAS number, substance name, SMILES, functional group (based on ECOSAR classifications), water solubility estimates, and log K_{ow} estimates [37]. The *test conditions* includes mean measured water concentration, radiolabeled substance (i.e. yes or no), exposure duration, temperature, pH, dissolved oxygen concentration, organic carbon content, lipid fraction, calculation method (i.e. steady-state or kinetic approach), combination exposure (i.e. exposure to a single substance or to a mixture), species at subcohort level, full-grown organism size (i.e. below or above 10cm [130]) and life stage. In addition, *endpoint information* includes the BCF values.

5.2.3 Data analysis

We used R [58] and the *nlme* package [131] to perform a linear mixed effects analysis of the relationship between the response variable (BCF value) and several test characteristics as fixed effects, including combination exposure, calculation method, species at subcohort level, life stage and full-grown organism size. The BCF values were lipid-normalized as advised in the OECD TG 305 and were log-transformed as standard deviation (SD) was correlated with BCF values (Figure S1). We used substances within functional classes as a random intercept in order to account for substance dependent differences within a functional class. This means that (average) BCF values are expected to differ per substance and that substances from the same functional class are expected to behave more similar than substances from a dissimilar functional class.

A three-step approach was followed. First, correlations between all fixed effects were investigated using bias-corrected Cramer's V. Of the five included variables, full-grown organism size was correlated (bias-corrected Cramer's $V > 0.7$) with life stage and combination exposure (i.e. single or mixture exposure) – and was excluded from further analysis.

Secondly, a candidate model set was constructed consisting of all possible additive combinations of fixed effects. Models with homoscedastic variances and heteroscedastic variances of the different fixed effects were included using the *varIdent* function. One model (full fixed effects and heteroscedastic variances for combination exposure, organism subcohort and calculation method) could not be run due to singularities. No interaction effects were included because of rank deficiency. All models were compared using the corrected Akaike Information Criterion (AICc), ranging from the null model (without any fixed effects) to the full model (including

all fixed effects). All models in the candidate set were fitted and then compared using AICc to determine the Kullback-Leibler (KL) best model [132]. The KL best model is the most parsimonious model (best fit to the data for the least number of parameters) given the model set. Additional models were considered to receive substantial support if the difference between model i AICc value and that of the KL best model (Δ AICc i) was < 2 [132].

Thirdly, we analyzed the contributions of the fixed effects on the means and SD for the best model and calculated marginal- and conditional-R^2. The marginal-R^2 describes the proportion of variance explained by the fixed effects and the conditional-R^2 describes the proportion of variance explained by both, the fixed and random effects. Visual inspection of residual plot of the best model did not reveal any obvious deviations of model assumptions. For relevant fixed effects, differences of the means were investigated using Tukey or Dunnett test for statistical variances.

5.3 Results

5.3.1 Included data

In total, 326 BCF values of 64 substances were included (details are given in Table S1). The BCF values ranged from log BCF 0.75 to 4.49 (i.e. BCFs ranging from 5.6 to 30625; data normalized to 5% lipid content) and estimated log K_{ow} values ranged from -2.15 to 6.79. For most substances three BCF values were included, though for some substances up to 23 BCF values were available. On average, two different references reported BCF values per included substance, with a maximum of six different studies. The substances covered eleven different functional groups (Figure 5.1A).

Different test conditions were applied in the included BCF studies (Figure 5.1B). Most BCF values were derived by a steady-state approach (n=299), whereas some were based on kinetic determinations (n=27). In addition, 149 BCF values were derived upon single substance exposure. Mixture exposures could be divided into organophosphate pesticides, halogenated organics and mixtures of polycyclic aromatic hydrocarbons (PAHs). Within the group of PAHs, two studies were included in which fish were exposed in combination with a potent mixed function oxygenase (MFO) stimulator, that mimics the metabolism induction of heavy weight PAHs (e.g. β-naphthaflavone) [133]. Furthermore, 17 different fish species were included, which could be divided into three groups based on subcohort level. The groups include the Neoteleostei, Protacanthopterygii and Ostariophysi (Figure S2). Common life stages include juveniles (n=187) and adults (n=114), though some studies used egg and/or larval stages (n=25). In addition, a clear balance was observed in the number of small (<10cm) and large species (>10cm), when considering their full-grown size. Within different

experiments, different combinations of test conditions were applied.

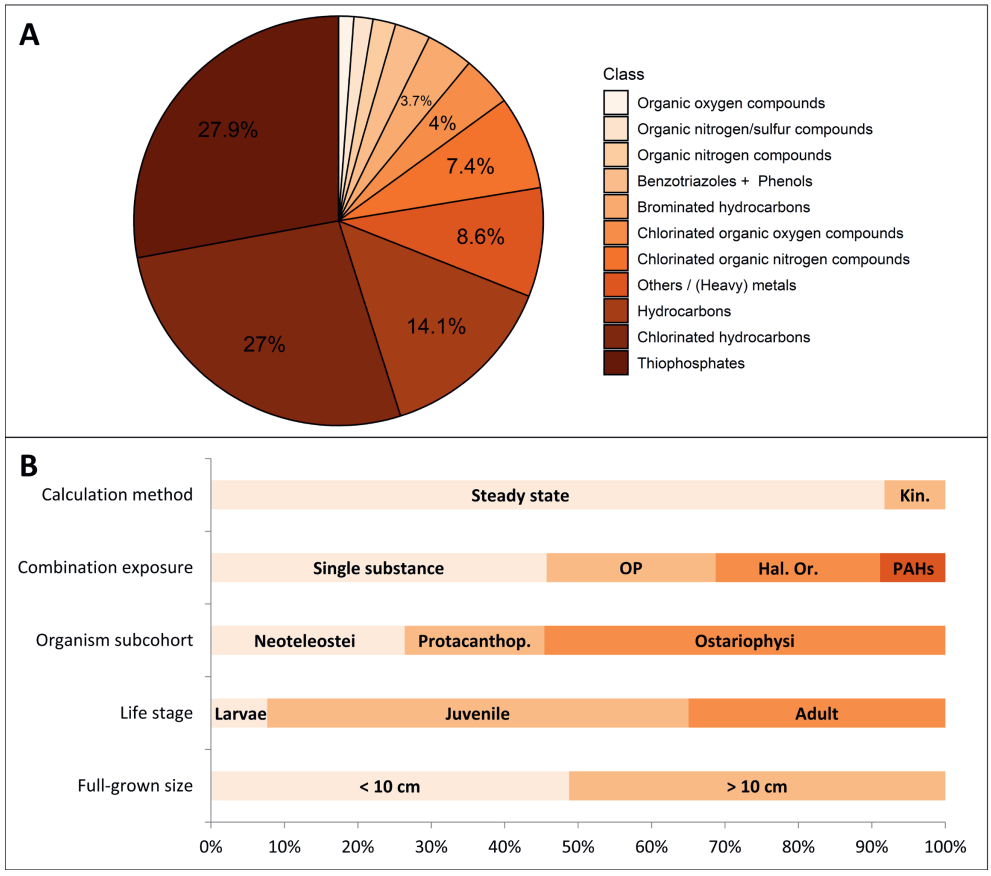

Figure 5.1: Overview of the included substances and BCF values. A) Overview of the functional classes of the different substances (n=64). B) Overview of the presence of different test conditions within the included test data (n=326). Kin. = Kinetic determination; OP = Organophosphate pesticides; Hal. Or. = Halogenated organics; PAHs = Polycyclic aromatic hydrocarbons (PAHs).

5.3.2 Explaining factors within BCF model

The results of the ten best descriptive models for bioaccumulation potential, based on AICc, are shown in Table 5.2. The top-ranked model included combination exposure, calculation method, organism subcohort, and life stage as fixed effects. Effectively, this means that those variables influence the BCF value. Furthermore, this model includes heteroscedastic variances for combination exposure, organism subcohort and life stage. Accordingly, differences in BCF variation (i.e. SD) are observed for different combinations of these variables. The top-ranked model had a marginal and conditional R^2 of 0.0974 and 0.843, respectively. Below, we discuss, for the top-ranked model, the differential effects of the included test characteristics

on obtained BCF values, and their variance.

Table 5.2: Overview of the top ten descriptive models. AICc = corrected Akaike Information Criterion. An "x" indicates the inclusion of a specific fixed effect within the model, or an allowance for heteroscedastic variances. Marginal and conditional R^2 are 0.0974 and 0.843, respectively, for the top-ranked model.

Rank	Fixed effects				Heteroscedastic variances				AICc	Δ AICc
	Combination exposure	Calculation method	Organism subcohort	Life stage	Combination exposure	Calculation method	Organism subcohort	Life stage		
1	x	x	x	x	x		x	x	282.5	0
2	x	x	x	x	x	x	x	x	285.3	2.8
3	x	x	x	x	x	x		x	290.3	7.8
4	x	x	x	x	x			x	294.6	12.1
5	x	x		x	x	x		x	294.6	12.1
6	x		x	x	x		x	x	296.4	13.9
7	x	x		x	x			x	297.9	15.3
8	x	x	x	x		x	x	x	300.5	18.0
9	x	x	x		x	x	x		303.0	20.5
10	x			x	x			x	307.9	25.4
…	…	…	…	…	…	…	…	…	…	…
80									364.9	82.4

Factors influencing bioconcentration

The set of test conditions were found to contribute differently to the BCF values. No difference in BCF value was observed when fish were exposed to a single substance or in a mixture with organophosphate pesticides or halogenated organics (Table 5.3). Substances that were tested in such mixtures were in general of the same class (i.e. organophosphate pesticides or halogenated organics, respectively). However, a significantly lower log BCF of 0.81 was observed upon exposure to a mixture of PAHs ($p < 0.0001$; Table 5.3). Further investigation revealed that PAHs were mainly tested simultaneously in combination with hydrocarbons and only ones in combination with an organic oxygen compound. For four substances, BCFs in our database had been generated upon single substance exposure as well as upon

exposure to a mixture of PAHs (Figure 5.2; including anthracene, dibenzofuran, fluorene and phenanthrene). From these substances it can be observed that the BCFs of three-ring PAHs (anthracene and phenanthrene) are much lower in case of exposure to a mixture of PAHs, whereas a small increase in BCF is observed in case of mixed exposure for the other substances.

Table 5.3: The effects on log BCF values of the test conditions that are included within the best descriptive model. Statistical analysis includes either Dunnett's test for combination exposure or Tukey's test for the other categories.

Group	**Comparison**		
	Compared to	**Effect in log BCF [SE]**	**p-value**
Halogenated organics	Single substance	0.07 [0.06]	0.2754
Organophosphate pesticides	Single substance	-0.11 [0.09]	0.2178
PAHs	Single substance	-0.81 [0.12]	*<0.0001*
Kinetic	Steady state	-0.27 [0.06]	*<0.0001*
Neoteleostei	Ostariophysi	0.15 [0.04]	*0.0003*
Ostariophysi	Protacanthopterygii	-0.17 [0.06]	*0.0022*
Protacanthopterygii	Neoteleostei	0.02 [0.05]	0.7767
Egg/larval stage	Juvenile stage	-0.08 [0.06]	0.2010
Juvenile stage	Adult stage	-0.02 [0.07]	0.7973
Adult stage	Egg/larval stage	0.10 [0.09]	0.2628

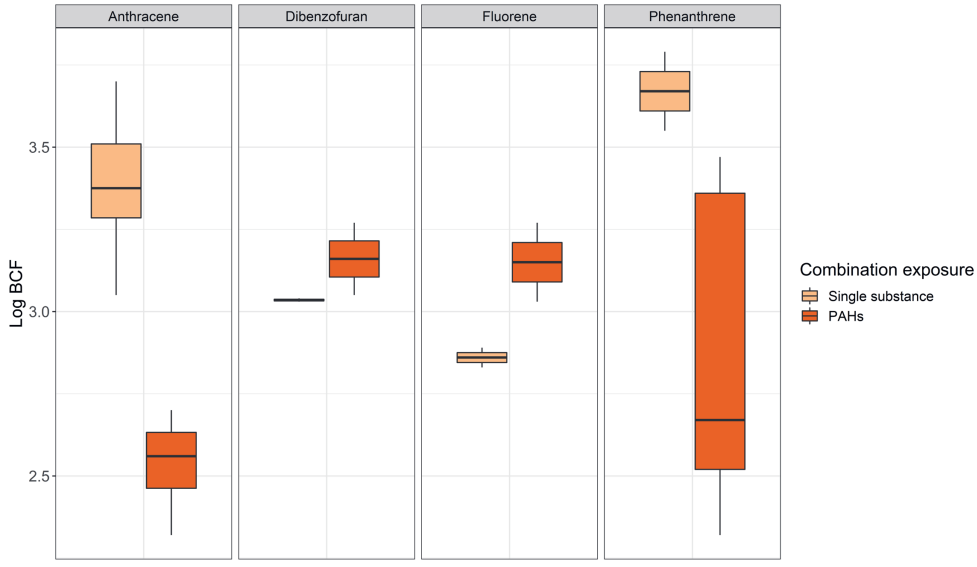

Figure 5.2: Overview of log BCF values for substances tested upon single substance exposure or upon exposure to a mixture of PAHs.

Furthermore, for 14 substances BCFs were determined via steady-state assessment as well as on the basis of kinetic approaches. The results indicate a significantly lower log BCF value of 0.27 when kinetically determined ($p < 0.0001$; Table 5.3).

The impact of subcohorts in the tests was found to result in different BCF values between organisms from the Ostariophysi as compared to the Neoteleostei and Protacanthopterygii subcohorts. For the group of Ostariophysi, which is mainly represented by the common carp (n=97; Figure S2), a lower log BCF of approximately 0.16 was observed ($p < 0.005$; Table 5.3). The Neoteleostei and Protacanthopterygii, which are mainly represented by the guppy and high-eyes medaka (n=37 and 23), and the rainbow trout (n=60), respectively, showed to have higher log BCF values.

Finally, life stage explains a certain amount of the variation in the data, as it is included as fixed effect within the top-ranked model. Lower BCF values are observed for egg and larval stages, compared to higher BCF values for adult fish (Table 5.3). However, no statistically significant differences of mean BCFs were observed between different life stages.

Variability in bioconcentration

Besides the influence of the test characteristics on the mean BCF values, also influences on SDs were estimated for different combinations of these characteristics. Within Table 5.4 all SDs are presented for groups of test characteristics with at least ten BCF values, which were corrected for dependent substance differences within functional classes. To clarify, when a substance will be tested multiple times in a BCF test using the following conditions: i) single substance exposure, ii) in an organism from the Neoteleostei subcohort, iii) at a juvenile life stage; a SD of 0.238 log BCF is expected to be observed based on available data. The observed SDs range from 0.090 to 0.343 log BCF with an average of 0.214 SD. To illustrate the average variation, 95% confidence ranges have been calculated in Table 5.5 for several BCF values, as based on 1.96 SDs of the mean.

5.4 Discussion

The aqueous exposure bioconcentration test is highly important for bioaccumulation assessments within regulatory frameworks. Nevertheless, little is known about the reproducibility and the factors within these laboratory experiments that affect the actual BCF value. Based on secondary data gathered within our database, we showed considerable impact of experimental design on the obtained BCF values and their variation. Specifically, mixture exposure, calculation method and the selected test fish species influenced the BCF values.

Table 5.4: SDs as calculated for combinations of test conditions that are considered relevant within the best descriptive model. Only groups of substances for which ten or more BCF values were available are included.

Combination exposure	Organism subcohort	Life stage	Number of BCF values	SD
Single substance	Neoteleostei	Juvenile stage	16	0.238
Single substance	Ostariophysi	Egg/Larval stage	12	0.090
Single substance	Ostariophysi	Juvenile stage	99	0.343
Halogenated organics	Neoteleostei	Adult stage	12	0.310
Halogenated organics	Protacanthopterygii	Juvenile stage	50	0.199
Organophosphate Pesticides	Neoteleostei	Adult stage	44	0.221
Organophosphate Pesticides	Ostariophysi	Adult stage	28	0.166
Organophosphate Pesticides	Ostariophysi	Adult stage	16	0.144

Table 5.5: The 95% confidence ranges of several BCF values based on the average SD of 0.214 log BCF.

BCF	Log BCF	Range ± 2xSD
100	2	38 – 262
500	2.7	191 – 1312
2000	3.3	762 – 5249
5000	3.7	1905 – 13122
10000	4	3810 – 26244

5.4.1 Influencing factors

Mixtures

A significantly lower log BCF of 0.81 was observed when the test substance was co-exposed with 4- or 5-ring PAHs. This was specifically observed for the 3-ring PAHs anthracene and phenanthrene (Figure 5.2). Earlier research indicated that single exposure to 3-ring PAHs did not stimulate the MFO system, whereas it was stimulated in combination with 4 or 5-ring PAHs [133]. Specifically, the MFO systems aryl hydrocarbon hydroxylase (AHH) and aniline hydroxylase (AH), as well as cytochrome P450 levels were induced by high molecular weight PAHs, including pyrene, chrysene and benzo(a)pyrene [133]. The MFO system is known to metabolize aromatic hydrocarbons by oxygenation and does not only act on higher weight PAHs. Consequently, lower BCF values are observed for 3-ring PAHs within a mixture of higher weight PAHs. Although no specific contributions were identified for mixtures containing organophosphate pesticides or halogenated organics, these findings suggest that results of mixture experiments should be interpreted with caution.

Calculation method

BCF values calculated based on kinetics resulted in lower log BCF values than determined by steady-state analysis. In theory, both approaches should provide similar results when uptake follows first-order kinetics and when steady-state BCFs are really based on steady-state data [19]. As it might be uncertain whether steady state is reached – especially for hydrophobic substances – kinetic BCF values are generally preferred [19]. If steady-state levels would not be achieved, one would expect to observe a lower BCF value for steady-state determinations. Nonetheless, we observed the opposite.

Potentially, the observed difference could be explained by a peak in fish concentration prior to achieving plateau levels. Such a phenomenon is regularly observed, and could be related to an interactive relationship between bioaccumulation kinetics and metabolic enzyme activities [134]. When a steady-state BCF is determined within this peak, a higher BCF value might be obtained compared to kinetic BCFs (Figure S3).

Organism subcohort

Data analysis revealed a significant difference in BCF values for species from varying subcohorts, with lower values for species from the Ostariophysi. This effect is likely related to differences in toxicokinetics.

The uptake of chemicals via the gills is generally related to the ventilation rate and the uptake efficiency [135]. The ventilation rate is described as the amount of water per time unit that is ventilated through the gills. The ventilation rate may differ across species, with higher rates for more active species [135]. The uptake efficiency, in the form of blood-water partitioning, is not assumed to vary between species for substances with a log K_{ow} above 3 [135,136].

Differences in depuration could be related to variances in metabolic activity among species, due to the presence of different biotransformation enzymes. Although many of those enzymes are very much conserved, different isoenzymes have been identified within different fish species, including different cytochrome P450 enzymes, glutathione S-transferases and ABC-transporters [137,138]. The presence and absence of many of those isoenzymes are related to the phylogeny of the species, and the activity of isoenzymes is thus likely to vary between different subcohort levels. As a consequence, varying V_{max} (i.e. the maximum reaction rate at saturating substrate concentration) and K_m levels (i.e. the substrate concentration at which the reaction rate is half of V_{max}) can be observed for different species [139,140]. For instance, differences have been observed within the metabolism of methoxychlor by the rainbow trout and the common carp, showing different metabolic profiles [141]. Only one metabolite was observed within rainbow trouts, whereas several metabolites were identified within carps. Despite information on the presence of different isoenzymes among (classical) fish species,

we lack knowledge on complex metabolic pathways of many substances and species. Better insight in these processes is considered valuable for risk management to quantify the variation across species.

Life stage

No significant effect of life stage on BCF values was seen, although a tendency of lower BCF values for the egg/larval stage, followed by juveniles and adult fishes was observed. Potentially, lower BCF values can be observed for early-life stages due to a larger growth capacity, resulting in growth dilution [19]. Furthermore, earlier research suggests that different life stages have different metabolic capacity, with varying V_{max} and K_m values [139]. However, also comparable differences in uptake rates have been observed [142], potentially resulting in comparable BCF values across life stages. Because of the comparable outcomes across life stages, the use of egg/larval stages might become of future interest to replace the standard *in vivo* bioconcentration test with non-protected *in vivo* systems [124].

5.4.2 Variability in bioconcentration

When considering the contribution of the different fixed effects, an average SD of 0.214 log BCF was determined. This variation is in line with the results of the OECD ring-test as conducted in 1985 by Kristensen and Nyholm [143]. Within this study, lindane was analyzed by 12 different laboratories testing one or two concentrations, resulting in a total of 22 BCF values. In addition, an optional chemical, 2,3,4,5-tetrachlorophenol (TeCP), was analyzed by four different laboratories, with in total seven BCF values. When normalizing the results to 5% lipid content, and only including the data that met the quality criteria of < 20% fluctuation in water concentration, a SD of 0.20 log BCF can be derived for lindane (n=19). For TeCP no reliable data could be retrieved according to the report [143]. The SD of 0.20 log BCF values, as derived under very strict conditions, is similar to our results.

While the above described test characteristics influence the BCF values, the remaining variation of 0.214 SD can be explained by other variables that were not yet considered in our analysis. Several factors have been suggested to potentially influence bioconcentration, including water-to-fish ratios [143], temperature [139,144,145], sex differences [139,144,145], feeding procedure (i.e. food item, feeding rate and feeding quantity) [146,147], and slight experimental variances in water chemistry and dissolved oxygen concentrations [148]. Most of these variables are expected to (in)directly influence the metabolic capacity of the organisms, and/or are directly related to changes in activity and oxygen consumption [149]. Indirectly, some of those factors might be partially covered by the inclusion of subcohort levels within the analysis. However, we can currently only speculate on the relative importance of all these variables, as many of them are not (consistently) reported. In addition, growth

dilution is known to significantly influence bioconcentration, especially for substances with a high bioaccumulation potential and for test organisms at early-life stages [118,150]. However, this parameter is scarcely reported and was therefore not included in the analysis. Moreover, part of the variability could potentially be related to variances in exposure concentration. As in theory the BCF is a net result of uptake and elimination rates, which are independent of exposure concentration [125], we did not consider this factor in the current analysis. However, concentration dependent BCFs could be of potential importance, specifically for polar chemicals, or for chemicals that undergo metabolic conversion when internal threshold concentration are attained [118].

In addition, it is expected that a significant amount of variation is related to intra-species differences. For instance, a two to three-fold variation is typically observed in the standard as well as maximum metabolic rate between individuals of the same fish species [151]. Individual differences are likely related to differences in genes and developmental conditions [151]. This may result in biological differences, like individual differences in isoenzyme content [139], and/or differences in behavior, like aggressiveness, boldness and (spontaneous) activity [151,152]. These factors are known to influence metabolic rates within organisms and subsequently affect ventilation rates, and thus may influence bioconcentration. A more accurate mean BCF (less influenced by the effect of individual differences) can be obtained by analyzing explicitly the biological variation within test organisms or by pooling or taking the mean of more samples [143], though sampling bias, due to behavioral differences, should be considered [153].

Besides the factors mentioned above, variation and uncertainty could also be related to laboratory practices, like fish maintenance, chemical analysis and data reporting. For instance, inadequate removal of uneaten food and/or feces may result in significant levels of organic carbon, limiting the bioavailability of the test substance [118,125,154]. Also differences in the analytical techniques (measuring chemical concentrations in water and fish), can contribute to the variation. Although it is generally assumed that the analytical methods are sufficiently optimized, variation may especially be observed for substances with a low water solubility. Moreover, we currently assumed that the selected water quality criteria (i.e. organic carbon, pH, temperature and dissolved oxygen concentration) were sufficiently strict to guarantee a limited influence on the BCF variability. Although some studies reported a range of water quality parameters that only partially met the criteria (n=5; see Table S1), exclusion of these values did not resulted in any changes on effect directions and significance levels. Nevertheless, also multiple studies did not report one or several water quality parameters and – following our approach – were included in the data analysis. This interpretation is a potential source of uncertainty, as extreme values for water quality parameters could significantly influence BCF variability [125]. We therefore encourage to report the water quality parameters in detail in

future studies.

5.4.3 Consequences for regulation and recommendations

When converting the SD to a 95% confidence range, an uncertainty of ± 0.419 log BCF is obtained (i.e. 1.96xSD; Table 5.5). This variation questions the robustness of the current B-assessment within regulatory frameworks, in which a single BCF value is generally sufficient to derive a conclusion. For example, a BCF value of 2500, which is normally interpreted as 'bioaccumulative', could also be considered as 'not bioaccumulative' and 'very bioaccumulative' based on the 95% confidence range (953-6561). The use of multiple experiments and/or species would be valuable for the B-assessment. Including more studies in order to encapture variability, has also been suggested for sediment quality assessments [155]. Potentially, new alternative bioconcentration methods based on invertebrate *in vivo* experiments could be valuable within such assessment, as they are less expensive and time consuming, and do not consider vertebrate testing [123]. The test performance of such methodologies could be compared and evaluated in the light of the performance of the current gold test standard as analyzed within this study (i.e. the aquatic exposure fish bioconcentration test). Specifically the use of alternative – non-vertebrate – bioconcentration tests should be stimulated, in order to further support the 3R principles (i.e. replacement, reduction and refinement of animal studies) [156,157]. Furthermore, we highlight that future studies should explicitly state and justify all experimental decisions and conditions, specifically also with respect to species-selection and simultaneous testing of substances. This is key, to improve the number of valid BCFs in databases.

5.5 Conclusions

Although guidance documents on bioaccumulation studies exist for many years and many studies have been performed accordingly, a review on reproducibility was lacking. Nonetheless, there is a crucial role of bioaccumulation assessment within regulatory frameworks. Our assessment indicates that several factors are influencing the bioconcentration potential, each of which should preferably be considered when interpreting the test results. The robustness of an experimentally determined bioaccumulation potential – although following the strict guidelines – is less than expected. We revealed a high variation in BCF values, with an average SD of 0.214 log BCF, within the fish bioconcentration test. Species selection and test designs where multiple substances are tested simultaneously showed to be important aspects leading to variation. The typical variability within BCF values results in high uncertainty in the B-assessment within regulatory frameworks. We, therefore, recommend the use of test species from at least two different subcohorts, including vertebrates or invertebrates.

Acknowledgements

This work was partially funded by the Dutch Ministry of Infrastructure and Water Management.

Supplemental material

Supplementary data to this chapter can be found online at https://doi.org/10.1016/j.chemosphere.2019.124731.