Journal of Hazardous Materials 501 (2026) 140648

Contents lists available at ScienceDirect

Journal of Hazardous Materials

journal homepage: www.elsevier.com/locate/jhazmat

# By integrating previously overlooked drivers AI boosts bioaccumulation assessment in fish

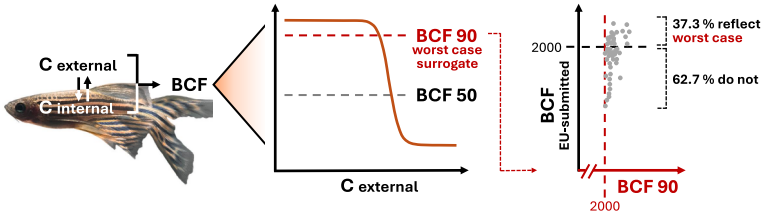
Heinz-R. Köhler [a,*,1], Reza Aalizadeh [b,c,*,1], Gabriele Treu [d,1], Thomas Gräff [d], Katharina Peschke [a,2], Ines Prutz [d], Nikolaos S. Thomaidis [c], Rita Triebskorn [a,e], Peter C. von der Ohe [d,**]

[a] Animal Physiological Ecology, University of Tübingen, Tübingen D-72076, Germany
[b] Department of Environmental Health Sciences, Yale School of Public Health, Yale University, New Haven, CT 06510, United States
[c] Laboratory of Analytical Chemistry, Department of Chemistry, National and Kapodistrian University of Athens, Panepistimiopolis Zografou, Athens 15771, Greece
[d] German Environment Agency, Section Chemical Safety, Dessau-Roßlau D-06844, Germany
[e] Steinbeis-Transfer Center for Ecotoxicology and Ecophysiology, Rottenburg D-72108, Germany

## HIGHLIGHTS

- The bioconcentration factor is not a universally valid measure for a chemical.
- High test concentrations result in low BCFs, low test concentrations in high BCFs.
- The BCF can be predicted with 90 % accuracy using an AI-derived application.
- A worst-case surrogate for chemical regulation is proposed (90 % of max. BCF, BCF90).
- About half of EU-submitted BCF data does not fulfil the BCF90 worst-case criterion.

## GRAPHICAL ABSTRACT




## ARTICLE INFO

*Keywords:*
BCF
Bioaccumulation
Fish
Machine learning
Organics
Registration

## ABSTRACT

The increasing use of chemicals has led to the integration of the bioconcentration factor (BCF) into chemical regulation. A machine learning model trained on chemical properties and test conditions for 962 chemicals enabled us to estimate experimental BCFs in 16 fish species with up to 90 % accuracy. We showed that the BCF is not, as generally assumed, a fixed chemical-specific criterion, but increases with lower exposure concentration and longer exposure duration. A review of 165 regulatory studies submitted to EU authorities showed that about half of the chemicals that should have been classified as bioaccumulative according to worst-case criteria defined in our study were not so classified based on the experimental BCFs submitted. We therefore propose the implementation of two new objective metrics, BCF50 for minimum requirement bioaccumulation screening and BCF90 as a realistic worst-case surrogate, into regulation.

* Corresponding authors.
** Corresponding author at: Department of Environmental Health Sciences, Yale School of Public Health, Yale University, New Haven, CT 06510, United States.
*E-mail addresses:* heinz-r.koehler@uni-tuebingen.de (H.-R. Köhler), reza.aalizadeh@yale.edu, raalizadeh@chem.uoa.gr (R. Aalizadeh), peter.vonderohe@uba.de (P.C. von der Ohe).
[1] Authors contributing equally to this work.
[2] Present address: Büro am Fluss GmbH, D-73240 Wendlingen, Germany.

https://doi.org/10.1016/j.jhazmat.2025.140648
Received 1 July 2025; Received in revised form 30 October 2025; Accepted 28 November 2025
Available online 29 November 2025

## 1. Introduction

Bioaccumulation is the accumulation of chemicals in biota relative to their environment. Bioaccumulative compounds biomagnify in food chains, putting especially top predators and humans at increased risk of adverse health effects [1–5]. The regulatory metric used for the assessment of chemicals in the aquatic system is the bioconcentration factor (BCF), which is defined as the ratio between the concentration of a chemical in an aquatic organism (e.g. fish) and the concentration of that chemical in the surrounding water body [6]. The BCF has been determined for many fish species commonly used for regulatory testing, either at a presumed steady state or during the kinetic phase of bioaccumulation [7, 8]. In the regulatory context, the octanol/water partition coefficient ($P_{ow}$ or $K_{ow}$) is commonly used as a measure of the tendency of a chemical to accumulate in organisms [9], and a screening criterion to assess the need for subsequent *in vivo* bioaccumulation testing. A substance is classified as Bioaccumulative (B) or Very/Highly Bioaccumulative (vB) if an experimental BCF value from *in vivo* tests exceeds thresholds of 1000 (B: Japan, USA) / 2000 (B: Australia, EU) or 5000 (vB: Canada, China, EU, Japan, USA), respectively [9] (see also supplementary Table S1). Regulators currently face several major challenges in assessing the bioaccumulation potential of a chemical. Firstly, the growing rate of chemical production and market introduction already surpasses the rate at which bioaccumulation assessment can be conducted and reliably evaluated. Furthermore, there is a significant deficiency in regulatory data regarding bioaccumulation potential for a large proportion of chemicals currently in use. This data gap is particularly pronounced for substances regulated under the EU REACH framework, especially those with low production volumes or low $P_{ow}$. Additionally, a considerable number of reported BCF values fail to satisfy the requisite quality standards. Moreover, very different BCF values have been reported for identical chemicals, most likely due to variation in biotic and abiotic factors in the respective tests [7,8,10,11]. However, such factors are often not specified or considered in regulatory acceptable test designs because their exact contributions have not systematically been evaluated and therefore remain unknown. Understanding the factors that influence BCF in fish laboratory tests is therefore essential for accurate hazard and risk assessments and compliance with future regulations.

Computational machine learning approaches, such as deep-learning (DL) and regression decision trees (DT) are expected to play a pivotal role in developing sophisticated *in silico* models. For application in regulatory frameworks, these models must provide accurate and reliable predictions and ideally also mechanistic insights, while in parallel clarifying their scope, limitations or uncertainties [12]. Deep learning (DL) and other ML models have been successfully applied to the prediction of various molecular properties, including bioaccumulation assessment, using the available experimental bioaccumulation data, mainly in fish [13–16]. However, a systematic adaptation for characterization of the bioaccumulation potential is limited, and assessing the reliability of individual predictions generated by complex "black box" models is a significant challenge. This complexity raises the need for equally powerful interpretability methods to gain confidence in the model, enable reproducibility, understand its limitations, and support the user's knowledge and intuition in the process of property optimization and design[13,17,18]. On the contrary, current Quantitative Structure-Activity Relationship (QSAR) models often have limited applicability as they are exclusively based on selected physico-chemical properties of organic chemicals, notably log $P_{ow}$, and additionally incorporate features of their molecular structures of varying complexity [14,16,19,20]. Moreover, they are unable to reliably model different BCF values for one and the same chemical that are documented in the literature. The predictive capability of existing BCF models is also restrained by their reliance on a limited dataset of experimental studies. Although there are already a number of promising approaches to model BCFs in fish, based on various types of regression, neural networks or AI methods (see supplementary Table S2), most of them are limited to modelling up to a few hundred data points at most (e.g. [21]). Only a few studies are based on more than a thousand chemicals, however, with a coefficient of determination ($R^2$) for the training set of only 0.775 at most [22]. Two other studies used data from 1129 and 1724 chemicals but predicting the test datasets only resulted in $R^2$ values of 0.77 and 0.75, respectively [20,23]. A multilayer perceptron-based study involving 3240 chemicals achieved even less success in predicting the test data set accurately ($R^2 = 0.62$, [24]). All the available studies to date also have in common the fact that the number of fish species for which BCFs can be modelled is either quite small (up to four, [21]) or unknown. As a result, there is no comprehensive method that is capable of accurately predicting BCF values across a diverse array of fish species while simultaneously quantifying the relative impact of relevant abiotic and biotic factors on the bioaccumulation potential of organic chemicals in fish. In response to the aforementioned limitations, our study pursued three key objectives: [1] to develop a novel, combined DL and DT model based on a harmonized and robust dataset of bioconcentration factors and relevant test conditions determined in experimental *in vivo* studies with fish; [2] to develop novel metrics to establish standards for universal comparability of bioaccumulation data, particularly with respect to realistic worst-case scenarios, which will allow the identification of appropriate test conditions for the determination of bioaccumulation factors to assist industry and regulatory authorities in generating more robust and relevant bioaccumulation data prior to the use of a chemical or for its registration and [3] to identify systematic errors in current bioaccumulation assessments in Europe by comparing BCF values predicted by the model for worst-case scenarios with values derived from confidential registration data from EU authorities, handled by the German Environment Agency.

## 2. Methods

BCF data were collected from publicly available literature (e.g. [7, 25–27] and existing databases, such as the "Gold Standard Database" of CEFIC and the "CERI" database of the Chemicals Evaluation and Research Institute (CERI), Japan (see Datasets 1–3 in the supplementary information (SI) for sources) for model training and validation. Canonical SMILES (Simplified Molecular Input Line Entry System) were compiled together with experimental and predicted endpoints (log$P_{ow}$, logS, $LC_{50}$ (*Pimephales promelas*), aerobic ready biodegradation and bioconcentration factors (*Cyprinus carpio*) using the ECOTOX module [28], TEST [29], OPERA [30] and data from open literature). The total dataset consisted of 7821 BCF values, of which 1709 had to be excluded due to them representing BCFs for specific tissues such as carcass, muscle or intestine rather than whole-organism measurements. Another 2320 data sets lacked essential information on test conditions or fish species. Thus, only 3792 studies had whole fish BCFs for known fish species, together with water concentration and exposure duration, of which about half related to steady-state studies in *Cyprinus carpio*. It should be noted that only 702 studies of them (18.5 %) reported lipid-corrected BCF values. In fact, no lipid-corrected BCF was reported in the literature on steady-state BCF studies for *C. carpio*. This restricted our ability to include this important factor in the current model.

A total of 10,286 different types of fingerprints (FPs), including standard FPs, graph FPs, circular FPs and chemical functional groups, were calculated for each SMILES entry using the 'rcdk v3.6' and 'checkmol' software packages for model development. A $500 \times 500$ pixel image of these FPs was then generated for each compound (see supplementary Figure S1). Pixels with an FP value of 1.0 were given a black colour (RGB colour code: 0, 0, 0), while those with a different value were left white (RGB colour code: 255, 255, 255). This pixel raster image was then correlated with the experimental values using the convolutional neural network (CNN) deep learning method with the TensorFlow package in Python. An example of the model architecture for regression and classification tasks is shown in Fig. S2. The size of the

convolution window (kernel), output channels (filters), convolution steps, activation functions, and hidden layers were optimised using a fitness function based on root mean square error (RMSE) from leave-one-out cross-validation analysis for each endpoint. The CNN-DL was trained on 75–85 % of the dataset (depending on the complexity and volume of the collected data) and tested independently using an external test set (15–25 %). Dataset partitioning was done randomly but was rigorously checked after model development to ensure that chemicals were selected fairly with respect to their chemical structure and experimental properties. Fairness was evaluated by checking the trend of the residuals against chemical similarity (k = 3 nearest neighbour chemicals using the Tanimoto index). In other words, chemicals showing a distinct trend due to diverse chemical structures were flagged and transferred to the training set to redefine the boundaries of the chemical space. This approach was also used to define the models' application domain. For a prediction for a new compound to be reliable, the trend of residuals of similar chemicals should decrease with increasing chemical similarity. However, an increasing trend in errors for more similar compounds could indicate an unexpected bias or error for a given compound. In this case, the predicted values would be unreliable. Further details can be found in the SI descriptions.

The model's application domain is hence defined by establishing the trend in residual change (predicted logBCF minus observed logBCF) over chemical similarity (Morgan fingerprints). First, the method calculates all pairwise similarity values; then, it filters out the lowest similarity values (k = 3) under the given similarity segments. Here, the chemical similarity segments are defined as follows: [1] above 0.9, [2] between 0.9 and 0.7, [3] between 0.4 and 0.7, [4] below 0.4. For each segment, the top three chemicals are retained, along with their information, including their residual values. Consequently, this method plots the mean residual changes as a function of the mean chemical similarity. To determine whether the error is due to chemical structural diversity rather than experimental error, the residual trend can provide a clue, as illustrated in supplementary Fig. S3. An acceptable trend — and, by extension, a reliable prediction — occurs when the prediction error decreases as the chemical similarity between the suspect and paired compounds increases. An unreliable prediction outcome occurs when the trend is not a function of chemical similarity and fluctuates regardless of the chemical similarity value. If the error increases as chemical similarity increases, there is an experimental error or residual that is not a function of chemical structure (Fig. S3). The latter type of trend requires more information or experimental validation to validate the prediction outcome. To further facilitate the recognition of trend patterns and the assignment of classes to each trend, we utilised a supervised classification technique based on a random forest (RF). Briefly, after constructing the similarity and residual data for pairs, each trend was investigated and assigned to one of the following three classes: reliable; acceptable (conditioned on further experimental analysis); or unreliable. The random forest then estimates the type of application domain (reliable, including the acceptable class, or unreliable prediction), bypassing the need for manual investigation of new suspect compounds and enabling fast screening as a function of high-throughput screening. The RF model was built using the 'randomForest' R package with the following parameters: 'proximity' was set to True and the 'mtry' parameter was set to 10 (the number of variables randomly sampled as candidates at each split to build each tree). The RF model was further tuned using the 'tuneRF' function in the 'randomForest' R package to avoid overfitting. Furthermore, MDSplot was applied to the developed RF to visualise the density of classes in the total database. In this case, the reliable versus unreliable chemical space could be determined and applied to a new case, as depicted in Fig. S3.

For external validation of the BCFpro model confidential data from 165 studies including metadata submitted to the European Chemicals Agency (ECHA), the European Food Safety Authority (EFSA) and the European Medicines Agency (EMA) that were handled by the German Environment Agency (UBA) for regulatory purposes were used and analyzed anonymously.

Statistical methods included ANOVA and multiple linear regression modeling. The machine learning approach was based on convolutional neural networks (CNN) and Extreme Gradient Boosting (XGBoost). Grey Wolf Optimizer [31] was used to evaluate the effect of experimental test conditions on the measured logBCF in 16 fish species. The experimental parameters considered included the BCF measurement approach (kinetic or steady-state), the duration of exposure (measured in days), and the nominal water concentration of the test substance (expressed in μg/L). The input variables for estimating logBCF values in the 16 fish species via XGBoost were the baseline logBCF value (from *Cyprinus carpio*), log $P_{ow}$ and log S (maximum water solubility), all predicted by CNN, fish species fingerprints and test conditions. The contribution of each of these variables in the final model (XGBoost) was calculated using Shapley values (SHAP). More details on the structure of CNN and the fine tuning of the XGBoost model can be found in the SI.

BCF data predicted by BCFpro for ionizable substances are based on neutral conditions (pH7, "initial" pH). For correlation analysis, we adjusted these predicted BCFs to other pH levels ("target" pH) where BCF values for these substances have experimentally been determined by [32] using the equation therein, however, adapted to the calculation of bioconcentration factors:

$$\log BCF_{\text{target pH}} = \log BCF_{\text{initial pH}} + (\log D_{\text{target pH}} - \log D_{\text{initial pH}}) \quad (1)$$

## 3. Results and discussion

Our research demonstrates that artificial intelligence significantly improves the computational estimation of BCF values in fishes as a function of influencing test parameters. The importance of an accurate and reproducible determination of the accumulation potential for the ecotoxicological classification within the environmental risk assessment of chemicals is underscored by the robust correlation observed between experimentally determined bioconcentration factors (BCFs) and the median lethal concentrations ($LC_{50}$) for both non-ionizable and ionizable organics ($P < 0.0001$; supplementary Fig. S4). Indeed, in many countries, substances are classified according to their potential to accumulate in biota by using log $P_{ow}$ as a measure of lipophilicity. However, our current meta-analysis reveals that solely relying on log $P_{ow}$ is insufficient for comprehensive bioaccumulation assessment. While the crucial role of log $P_{ow}$ and water solubility in determining a chemical's uptake and accumulation potential has been long-established and mechanistically elucidated [33,34], our results reveal that these parameters account for only 45.4 % and 45.2 % of the variability in reported BCFs, respectively. This finding emphasizes the necessity of considering a broader array of relevant parameters in conjunction to achieve accurate bioaccumulation predictions (see SI text).

### 3.1. Parameters that determine the BCF

Our study shows that the experimentally determined BCF for a given substance, such as that obtained during registration procedures, is not universally applicable independent of contextual parameters. This finding is particularly noteworthy, as often only a single, ostensibily reliable substance-specific BCF value is required for authorisation procedures. Contrary to regulatory assumption, our research shows that in addition to intrinsic properties such as the lipophilicity of the chemical and the accumulation capacity of the fish species under consideration (i.e. its fat content), the reported BCF in fish is decisively influenced by two experimentally variable parameters: the applied concentration of the chemical in the external medium and the exposure time (Fig. 1A). The assumption that the observed dependence of the BCF on the test concentration only occurred because predominantly lipophilic chemicals with lower water solubility were tested at low concentrations and –

because of their limited water solubility – none of them at high concentrations, thus resulting in high BCFs due to their lipophilicity and, therefore, creating a pseudo-correlation, could be clearly refuted. For all orders of magnitude of log $P_{ow}$ between −3 and 10, we demonstrated this negative correlation between test concentration and BCF separately (see Fig. 1B). This correlation was also highly significant in narrow log $P_{ow}$ ranges with high data density where log $P_{ow}$ varied by a maximum of only 0.05. Examples include log $P_{ow}$ 2.84–2.88 ($n = 37$, $p < 0.0001$), log $P_{ow}$ 3.98–4.02 ($n = 131$, $p < 0.0001$), log $P_{ow}$ 4.55–4.60 ($n = 128$, $p < 0.0001$), log $P_{ow}$ 5.10–5.15 ($n = 137$, $p < 0.0001$) and log $P_{ow}$ 5.92–5.95 ($n = 129$, $p = 0.0004$). Thus, in this context, our study reveals that the variability in BCF values as a function of external chemical concentration is comparable to the magnitude of variability observed when BCF is plotted against lipophilicity. This is a particularly significant finding as lipophilicity (i.e., log $P_{ow}$) is ***the*** parameter commonly used for bioaccumulation estimation, based on the well-known inverse relationship between water solubility of a chemical and lipid content in fish [35, 36].

Variations in BCF of more than two orders of magnitude are permissible for any given chemical which, according to current OECD practice, leaves a wide range in the choice of exposure conditions to achieve intended BCF values. For this reason, we advocate for the development of standardized, exposure-independent BCF metrics for the assessment of chemicals which objectively reflect the accumulation potential of chemicals in an environmentally relevant manner, thereby improving consistency and comparability across different studies and regulatory frameworks.

The OECD Test Guidance (TG) Document [8] on aspects of the OECD Test Guideline No. 305 on bioaccumulation in fish [21] erroneously assumes for the testing of "moderately hydrophobic compounds", based on only two studies [38, 39], that for the "vast majority of pesticides and general chemicals" the BCF is independent of the test concentration in water and therefore often only one test at an unspecified concentration is sufficient. Using 3792 data mainly arising from public sources (Fig. 1B) and for registration data submitted to the EU (Fig. S5), our study particularly revealed and quantified the extent of a general negative relationship between the water concentration of a chemical and the resulting BCF. This observation has been occasionally reported in previous studies also [40–45] but has never systematically been investigated. Although the OECD TG 305 attributes this effect to the absence of metabolic processes at low concentrations, our research suggest a different explanation. We propose that this phenomenon is an inherent consequence of a general mechanistic interrelationship between the internal aqueous and lipid phases in exposed fishes (Figs. 2, 3 and S6, detailed explanation in the SI) and therefore crucial for the importance of the 'chemical concentration' parameter in determining a BCF.

The established view [37,46,47] is that the BCF is, by definition, calculated as a steady-state ratio of the chemical concentration in fish to the chemical concentration in water: $BCF_{steady\ state} = C_{fish} / C_{water}$. Theoretically, the $BCF_{steady\ state}$ is calculated when there are no changes in concentration in the fish or the water over time. Alternatively, the BCF can be calculated as the ratio of the uptake rate constant from water ($k_1$) to the total elimination rate constant to water ($k_2$): $BCF_{kinetic} = k_1 / k_2$. According to this theory, there should be a single 'true' BCF for a given chemical and fish species that does not vary with exposure duration or concentration. However, this theoretical view is based on two assumptions. Firstly, it requires a steady state, which would necessitate lengthy experiments, particularly for highly lipophilic substances [48] and is therefore not always achievable. Secondly, it assumes that the

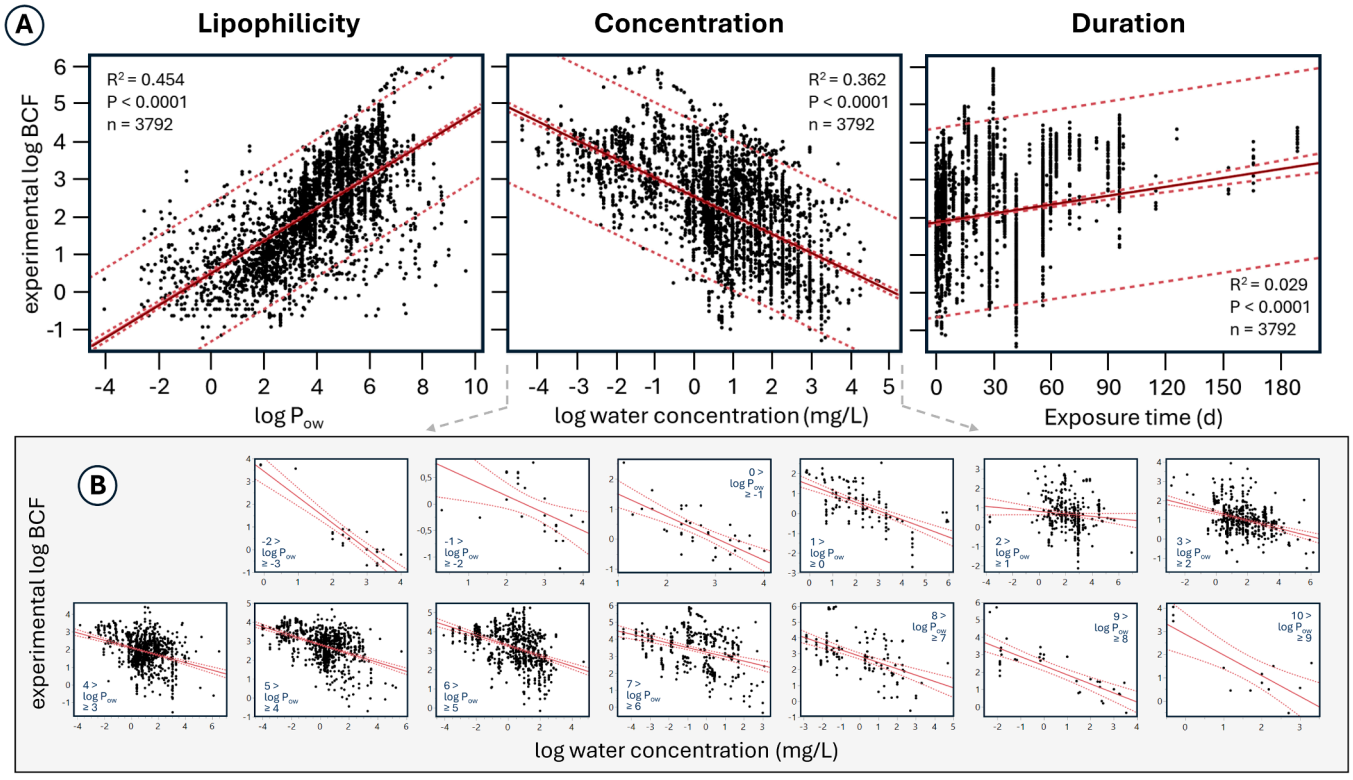



**Fig. 1.** (A) Experimentally generated log BCF data *vs.* log $P_{ow}$ of corresponding chemicals (left), their log concentration in the exposure water (middle) or the exposure time (right). Own compilation of literature-based data from public sources. Linear regression line (red, solid) and its 95 % confidence intervals (red dashed, inner interval) plus 95 % confidence intervals for the data point distribution (red dashed, outer interval). Highly significant ($P < 0.0001$) correlation (ANOVA) of the parameters in all cases. $R^2$: Bravais-Pearson variance. **(B)**: Separate linear regression and ANOVA analysis for subsets of the database, separated according to the order of magnitude of the log Pow of the chemicals. These subsets spanned 13 orders of magnitude, from −2 > log Pow ≥ -3–10 > log Pow ≥ 9. Significance at $P < 0.0001$ was observed in all cases except for −1 > log Pow ≥ -2 ($P = 0.009$), 1 > log Pow ≥ 2 ($P = 0.07$), and 10 > log Pow ≥ 9 ($P = 0.001$).

conditions for experimentally determining the BCF are the same as those for measuring chemical distribution in two phases, e.g. when determining the log$P_{ow}$ of a chemical. As we will show, this is not the case, and this has major consequences.

Of course, the BCF represents a distribution ratio between two phases (aquatic organism/surrounding water) that is, in principle, comparable to the $P_{ow}$ (organic phase/aqueous phase). The $P_{ow}$ is experimentally determined using equal volumes of aqueous and organic phases. The more a substance is present in the organic phase, the less it is present in the hydrophilic phase, resulting in a defined concentration ratio that is independent of the amount of chemical used. However, this principle does not apply to the distribution ratio of a chemical present in a given volume of water and its presence in an exposed fish, since the respective phases are not of equal volume and the uptake and accumulation of a chemical in a fish contributes negligibly little to reducing the concentration of that chemical in the much larger volume of surrounding water. This holds true under natural environmental conditions and laboratory settings for a majority of chemicals, specifically those with a log $P_{ow}$ less than six (see a sample calculation on page 5 of the supplementary text). Consequently, equilibrium will not be established between the proportion of the chemical absorbed by the fish and the external medium, since a constant concentration in the disproportionately larger external medium will continually cause 'new substance' to reach the fish's surface and passively 'push' its way in. This process will ultimately continue until the maximum solubility of the lipophilic substance in the fish's fat content is reached. Furthermore, bioaccumulation in fish goes beyond the simple partitioning of a compound between a hydrophilic and a lipophilic medium (as in $P_{ow}$ measurements), as both aqueous and lipid components are present in the fish body. In addition, the uptake and storage capacity of an organic compound in a fish is highly dependent on various parameters such as age, size and, most importantly, lipid content. These parameters also limit the maximum storage capacity of the chemical in the exposed organism, which therefore reaches a steady state after a certain period of time. This period is longer for low concentrations in the external medium than for higher concentrations.

While the concentrations of chemicals in the exposed organism increase over time approaching a steady state, their concentrations in the surrounding water remain essentially constant over time. Furthermore, the maximum storage capacity of the lipid fraction in fish is independent of the external concentration of a chemical (at lower concentrations, it simply takes longer to reach the maximum level of lipid solubility). Thus, as shown in Fig. 1 of our study, the resulting BCF, a ratio of the chemical's concentrations in the fish to that in the surrounding water ($C_f/C_w$,) is, in consequence, inversely related to the initial water concentration, because the denominator ($C_w$) remains essentially constant over time. This mathematical principle explains the observed negative concentration-BCF relationship (Figs. 3, S6 and SI text) and applies to all exposure situations in which the volume of the external medium greatly exceeds that of the test organism. Furthermore, the applicability of this principle extends beyond steady-state conditions at maximum internal chemical concentration but is also relevant in scenarios where only certain percentages of the maximum internal concentration are reached, e.g. during the kinetic phase of the accumulation process during testing (Fig. 2 and SI text). In this context, it is also mechanistically evident that exposure time plays a crucial role in the kinetics of accumulation of an organic chemical in biota as, at a constant concentration in the external medium, the saturation concentration determined by the lipid content in the animal (Fig. S7) is increasingly approached with increasing exposure time (Fig. 2). Consequently, also the BCF exhibits a corresponding increase over time. This simple relationship between lipid content and bioaccumulation underlies the practice of normalizing results to a "standard" lipid content of 5 %, to enhance comparability of results [8]. This relationship likely explains the variability in BCFs among fish species with different lipid contents. Thus, our study strongly contradicts the recent proposals to revise regulatory guidance, e.g. the EU Guideline for Environmental Risk Assessment of Medicinal Products for Human Use, to (a) increase the log $P_{ow}$ trigger value from > 3 to > 4 for requesting experimental studies, and (b) conduct BCF tests in fish only at single exposure concentrations arguing that BCFs are not affected by water concentrations [25].

The mechanistic relationships between the net uptake kinetics, considering both uptake and elimination of a given organic chemical at different exposure concentrations as a function of exposure time, and the consequences for the resulting BCF, are visualized in Fig. 2 (see SI for detailed explanations on the rate constants for uptake, $k_1$, and depuration, $k_2$; and on the way to calculate BCF = $k_1/k_2$) and Fig. 3 (see supplementary Fig. S6 and SI text for an example calculation of BCF = $C_f/C_w$).

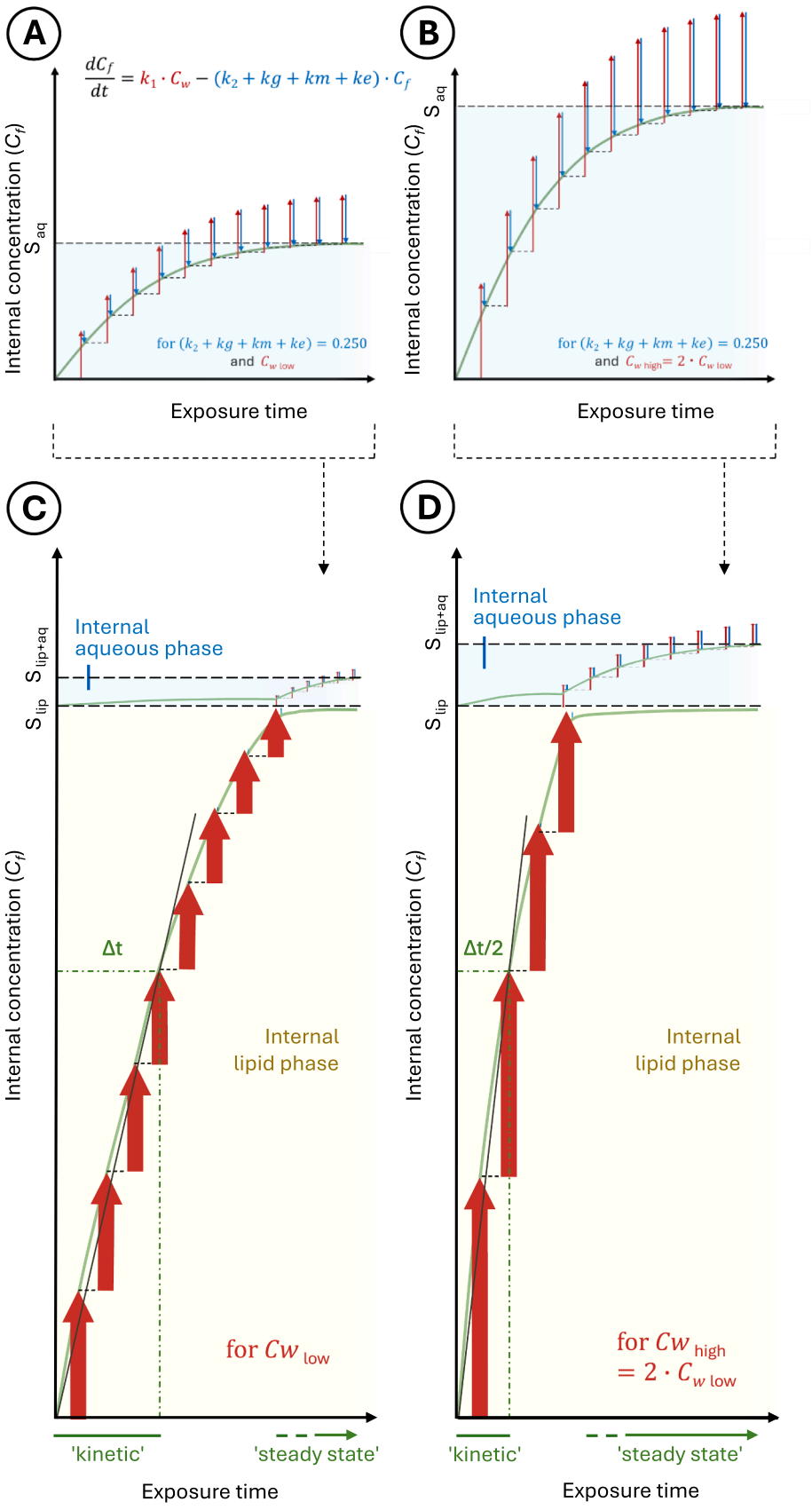
### 3.2. Fish as a two-phase (lipid-water) system

The lipid content of fish can vary considerably between individuals, species, and even within the same fish depending on factors such as season, diet, and reproductive state [25]. Studies in the aquatic environment have shown that there is a significant positive correlation between the accumulation of a chemical and the lipid content of organisms, so that residue levels vary between individuals, species, and size groups depending on their lipid content [49,50]. Therefore, fish exposed to very low aqueous concentrations of lipophilic chemicals can experience quite high internal concentrations (and high BCFs). As the fat content of an animal increases, the saturation concentration ("s" in Figs. 2 and 3) for the organism also increases. Therefore, at least under steady-state conditions, higher internal concentrations of a chemical are inevitable. Consequently, the BCF will also increase. Also, during the uptake kinetics, higher lipid contents at the same external concentration and test duration lead to higher uptake rates and thus also to higher BCF. This finding is very consistent with the general fish bioaccumulation model [51] according to equation (2) (Fig. S7, for explanation see figure caption).

$$\frac{dCf}{dt} = k_1.C_w - \left(k_2 + k_g + k_m + k_e\right).C_f \quad (2)$$

For example, a study on fenthion in fish showed that acceptable wet weight BCF values could vary by a factor of 3 or more depending on the lipid content of the fish [25]. OECD TG 305 requires that the BCF is corrected for a 5 % lipid content (for both steady- state BCF and kinetic BCF) during the study period to reduce variation in BCF values [8]. Lipid normalisation is critical for generating consistent, comparable, and meaningful bioconcentration data in fish studies, especially for lipophilic substances. It helps account for natural variability in fish lipid content that would otherwise confound the interpretation of bioaccumulation results. However, for the majority (>80 %) of the experimental data available, the lipid content of the fish was not reported. Therefore, the effect of fish lipid content could thus not be included in the BCFpro modelling, but should be addressed in further work as more data become available. A wide range of procedures for lipid extraction procedures are now available, but only appropriate techniques which allow complete extraction of total lipids should be selected in bioaccumulation studies to ensure the correct lipid normalisation of BCF values [52].

The following considerations are critical to understanding the importance of interactions between the internal lipid phase and the internal aqueous phase in fish for the BCF. According to OECD TG 305, it is possible to calculate the BCF = $k_1/k_2$ for "moderately lipophilic" substances by applying equation (2). Based on theoretical considerations, this is certainly possible for hydrophilic substances, since they basically are not supposed to interact with the internal lipid phase in the fish. For these substances, the BCF – by theory – should also be independent of the test concentration in water (at realistic test concentrations well below the maximum water solubility). As our analysis of multiple experimental datasets shows, this is not the case, even for compounds that are highly soluble in water (see Fig. 1B). This is eventually due to




(caption on next page)

**Fig. 2.** Theoretical considerations on the accumulation kinetics of a given organic chemical as a function of exposure time in fish exposed to low (A, C) and high (B, D) external water concentrations ($C_w$). According to the OECD 305 Test Guidance (TG) Document [8] the general fish bioaccumulation model is described by the uptake rate $k_1 \cdot C_w$ (in red) and the overall loss rate $(k_2 + k_g + k_m + k_e) \cdot C_f$ (in blue). Uptake and loss are displayed by arrows (in red and blue, respectively). The differential equation given in the TG refers to the internal aqueous phase, i.e. the aqueous medium in the cells of a fish and in its tissue fluids (A, B): $k_1$ = first order rate constant for uptake into fish ($L \cdot kg^{-1} \cdot d^{-1}$), $k_2$ = first order rate constant for depuration from fish ($d^{-1}$), $k_g$ = first order rate constant for fish growth ('growth dilution') ($d^{-1}$), $k_m$ = first order rate constant for metabolic transformation ($d^{-1}$), $k_e$ = first order rate constant for faecal egestion ($d^{-1}$), $C_w$ = concentration in the water ($mg \cdot L^{-1}$) and $C_f$ = concentration in fish ($mg \cdot kg^{-1}$ wet weight). (A) Exposure to a given concentration of a (moderately) lipophilic chemical, $s_{aq}$ denotes the saturation level in the internal aqueous phase of the fish. (B): Exposure to twice the concentration of the same chemical compared to (A) results in twice the internal concentrations ($C_f$) of the chemical, both during the kinetic phase of accumulation and at steady state ($s_{aq}$) according to the general fish accumulation model at constant $k_1$ and $k_2$. Considering only the internal aqueous phase of the fish, the BCF for a given chemical and fish species, whether calculated by $C_f / C_w$ or by $k_1 / k_2$, would not differ between situations (A) and (B) because in case (B) both $s_{aq}$ and $C_w$ would be double as high as in case (A), and thus the BCF would be virtually independent of the concentration in the surrounding water. However, our results in this study show that, for the most part, this is *de facto* not the case. In fact, the fish accumulation model does not consider the different conditions prevailing in the internal aqueous (blue-shaded) and the lipid (yellow-shaded) phases of the fish. (C) and (D) show the general relationships for these two. In the lower part of both graphs, the internal lipid phase (shaded in yellow) is shown, and in the upper part, the optically compressed conditions as shown in (A) and (B) (shaded in blue). The saturation of the lipid phase is labeled $s_{lip}$, the saturation of the whole fish is labeled $s_{lip+aq}$. (C): Low test concentration of the chemical in the water, (D): twice as high concentration of the same chemical. In both cases, note that for a chemical with $logP_{ow} = 5$ and a fish with 5 % fat, the lipid phase contains 5000 times more molecules of the chemical than the aqueous internal phase due to its $P_{ow}$-based distribution. In this lipid phase, however, the rate constant for depuration from the lipid phase ('k4' in equations [3] and [4]) is close to zero, since the only way the chemical can be removed from the fish is by diffusion back into the internal aqueous phase of the fish and then back again into the ambient water phase, which is very infrequent for lipophilic chemicals with a high $logP_{ow}$. Consequently, a lipophilic chemical in both case (C) and case (D) will accumulate mainly in the lipid phase (light green areas) - in case (D) twice as fast as in case (C) - until the maximum saturation level $s_{lip}$ is reached in the steady state. In the internal aqueous phase, the relationships described by the general bioaccumulation model apply according to (A) and (B), but the existing concentration differences (shaded in blue) in this aqueous phase are, due to their comparatively low importance in the distribution of lipophilic substances (at $logP_{ow} = 5$: ratio 1:5000, see above), almost irrelevant for the resulting total internal concentrations ($s_{lip+aq}$ in the steady state) and thus practically negligible (which has been transposed in Figs. 3 and S6 for reasons of comprehensibility). Only in the initial phase of accumulation (labeled "kinetic" on the x-axis) the BCF can be reliably determined by the quotient $k_1/k_2$ (black rising lines in (C) and, with double slope, in (D)). However, as the exposure period progresses, the fish are in a phase that corresponds neither to the classical 'kinetic' uptake phase nor to the final steady state condition. In this intermediate phase, the concentration of the chemical in the external medium begins to become increasingly important for the BCF. We assume that the experimental determination of the BCF of fish in very many studies takes place in this intermediate range, since the dependence of the determined BCF values on the external concentration of a chemical is statistically demonstrable (see Fig. 1). When calculating the BCF $= C_f / C_w$ for the steady state situations in (C) and (D), the resulting value for the identical substance in fish with identical fat content in (D) would therefore be about half as high as in (C), since the $C_w$ in the denominator in (D) is twice as high as in (C), and the $C_f$ in the numerator (which is close to $s_{lip+aq}$) hardly differs between the two scenarios.

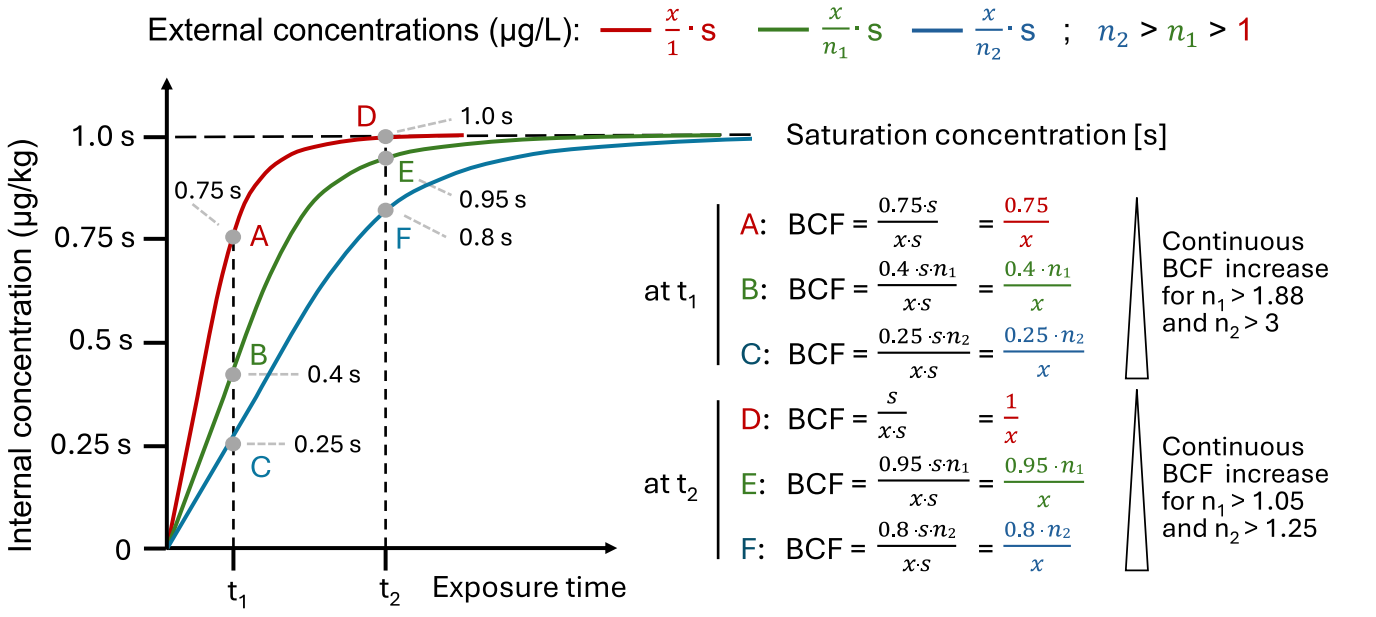



**Fig. 3.** Accumulation kinetics of a given organic chemical as a function of exposure time. Example plot for 3 different external concentrations (red, green, blue; concentration values as products of the saturation concentration, s, with the highest concentration in red (e.g. 1) and multiple times lower concentrations (e.g. $1/n_1$, $1/n_2$) in green and blue). Calculation of the resulting BCF values ($BCF = C_f/C_w$) for 6 arbitrarily selected points (A-F). It is evident that at both arbitrary times, $t_1$ and $t_2$, the BCF increases with decreasing external concentration, provided that at $t_1$ the mean concentration (green) is at least 1.88 times lower and the lowest concentration (blue) is at least 3 times lower than the highest concentration tested (red). For the later time point $t_2$, the condition for a continuous increase in BCF with decreasing external concentration is that the mean concentration (green) is at least 1.05 times lower and the lowest concentration (blue) is at least 1.25 times lower than the highest concentration tested (red). Realistically, this is the case for the vast majority of tested concentration series, commonly using a spacing factor of ten. It is also evident that increasing the exposure time leads to higher BCFs.

limited bioavailability of higher concentrations, caused by absorption effects or precipitation. However, with increasing $P_{ow}$, even a theoretical independence of the BCF from the concentration in the external aqueous medium is no longer given for more lipophilic chemicals with increasing exposure time. The higher the $P_{ow}$ of a chemical, the more it interacts with the internal lipid phase in the fish, so that the uptake kinetics - neglecting growth dilution, metabolic transformation and fecal egestion - are described increasingly poorly by equation (2), but rather by equations (3) (transition between external aqueous medium and internal aqueous phase) and (4) (transition between internal aqueous phase and internal lipid/fat phase):

$$\frac{dC_{fa}}{dt} = k_1\prime . C_w - k_2\prime . C_{fa} \quad (3)$$

$$\frac{dC_{fl}}{dt} = k_3 . C_{fa} - k_4 . C_{fl} \quad (4)$$

Here, $k_1\prime$ = first order rate constant for uptake into the fish´ internal aquatic phase, $k_2\prime$ = first order rate constant for depuration from the fish´ internal aquatic phase, $k_3$ = first order rate constant for uptake into the fish´ internal lipid/fat phase, $k_4$ = first order rate constant for depuration from the fish´ internal lipid/fat phase, $C_w$ = concentration in the external water, $C_{fa}$ = concentration in the fish´ internal aquatic phase, and $C_{fl}$ = concentration in the fish´ internal lipid/fat phase.

In equation (4), $k_3$ and $k_4$ depend crucially on the $P_{ow}$ of the chemical in question. For highly lipophilic substances, $k_3$ is very high (at least initially), while $k_4$ is comparatively low, according to the high $P_{ow}$. As shown in Fig. 2, in such cases the vast majority of a highly lipophilic chemical in fish is initially incorporated into the lipid phase (large red arrows in Fig. 2). However, with increasing saturation of the lipid phase, $k_3$ will continue to decrease because even lipophilic substances cannot reach arbitrarily high concentrations in the lipid phase of the fish. In parallel, $k_4$ increases with time until $k_3$ and $k_4$ become increasingly similar when approaching lipid phase saturation. The lipid phase will eventually reach its maximum saturation level, however, this saturation threshold cannot be surpassed.

Using the two-phase model presented here, the BCF can be calculated using equation [5] with BF (body fat) denoting the lipid proportion:

$$BCF = (1 - BF) . \frac{k_1\prime}{k_2\prime} + BF . \frac{k_3}{k_4} \quad (5)$$

The first term refers to the internal aqueous phase, the second term to the internal lipid phase. It should be noted that if the lipophilicity of a test chemical is high, $k_4$ is, accordingly, initially low and increases slightly when approaching the saturation of the lipid phase, and $k_3$ is initially high and decreases substantially with increasing exposure time – the higher the test concentration in the external medium, the sooner $k_3$ decreases. Therefore, it can be mathematically justified that the BCF can be reliably determined only after a short exposure time using equation (2) but will later continuously decrease with increasing test concentration and thus rapidly decreasing $k_3$ (with $k_1\prime$ and $k_2\prime$ remaining constant and a rather low but increasing value for $k_4$). Since $k_3$ and $k_4$ are not constant when approaching the lipid saturation level, the calculated BCF decreases as internal lipid saturation is approached – and the higher the test concentration of a lipophilic chemical in the external medium, the sooner this effect will take place. This is consistent with the observations in our study. Although $k_3$ and $k_4$ cannot be determined using the OECD TG 305 experimental approach, the relationships described by equation (5) are essential for understanding the underlying principles. In a hypothetical accumulation experiment – theoretically conceivable, but not practical – in which a lipophilic chemical would be applied in the external medium with maximum lipid solubility, the BCF can, according to these considerations, be a maximum of 1.0, but only if the fish would consist of lipid to 100 %. Conversely, even an extremely water-soluble chemical, tested at saturation in the aquarium water on fish with a negligibly low lipid content, would only accumulate with a maximum BCF of 1.0. These extreme examples illustrate that, also based on theoretical considerations, the BCF cannot be independent of the concentration tested.

### 3.3. Modelling and the BCFpro application

A comprehensive analysis of 3792 experimentally determined BCF values for 962 chemicals in fish, with known test conditions, revealed significant insights through multiple regression modeling. The best-fit model ($P < 0.0001$), incorporating chemical parameters such as log $P_{ow}$, log water concentration, log water solubility, exposure time, temperature, and fish species, accounted for just 64.9 % of the total variation in BCF values (Fig. S8; for some species, only less than 50 % of the total variation could be explained). While each of these parameters contributed significantly to this model, the remaining unexplained variance suggests the existence of yet unidentified interactions and interdependencies among these parameters. To further improve predictive capabilities, a deep learning-based model (based on convolutional neural networks) for *C. carpio* was used as input to the XGBoost regression tree. This final model hence incorporated additional intrinsic parameters, including chemical structure represented by SMILES notation to predict the BCF for unknown substances under variable exposure conditions such as different fish species, chemical concentrations, and exposure times, (Figs. S1-S2, S9-S10). When checking the quality of these estimates against the experimental data in our compiled datasets, the *in silico* predictions achieved an accuracy of up to 90 % when compared to experimental data, an exceptionally high concordance correlation coefficient of almost 1 and a RMSE of about 0.4 log units (Fig. 4, details on fish species in Figs. S11-S20). Even the correlation ($R^2$) between log BCF and log $LC_{50}$ improved by 9.4 % when the BCF predictions generated by deep learning were used instead of experimental data (compare Figs. S4 and S21). Compared to previously published models for BCF prediction, our combined DL and DT approach stands out due to its extensive data base, high level of accuracy, and exceptionally broad applicability (Fig. 5, supplementary Table S2 and SI text). The Grey Wolf Optimizer (GWO) [31] algorithm was subsequently employed to evaluate various combinations of exposure conditions to search for optimal parameters in BCF experiments, exploring a parameter space that included 16 different fish species, chemical concentrations ranging from 10 $pgL^{-1}$ to 100 $mgL^{-1}$ with a fine-grained slicing size of 1 $pgL^{-1}$, and exposure times varying from 2 h to 190 days with a slicing step of 1 h. The GWO algorithm can calculate the realistic situations for which the BCF would be maximised, using numerous combinations of exposure time and water concentration to predict the corresponding BCF (see Fig. 6, inlet 6: Optimisation of test conditions by Grey Wolf Optimization (GWO)). It can also calculate the maximum possible BCF and percentage shares of it (e.g. 90 % or 50 %). So, the objective of the GWO algorithm was to identify the specific exposure conditions (water concentration and duration) that would result in 90 % and 50 % of the maximum BCF achievable under realistic laboratory conditions. We define two substance-specific metrics, BCF90 and BCF50, which represent 90 % and 50 % of the maximum BCF respectively, integrating across various test conditions and fish species. Based on this XGBoost-GWO system, we have developed an application called 'BCFpro' (Fig. 6, freely accessible at https://www.parcopedia.eu/BCFpro) that determines reliable BCF values for unknown organic substances for 16 fish species, including the most commonly used test species worldwide, visualizes their expected variability and specifies the experimental conditions necessary for simulating realistic scenarios, such as BCF90 (representing a realistic worst-case scenario) or BCF50 (representing a minimum requirement scenario). Reliable BCF predictions are also given for ionisable chemicals (with 70 % explanatory probability, Fig. 4) whose BCFs depend on the ambient pH, taking into account the recently introduced mathematical variable ΔlogD [32], i.e. the difference of the lipophilicity of ionisable chemicals at two different pHs. ΔlogD has already been used to set a stricter EU EQS for ibuprofen

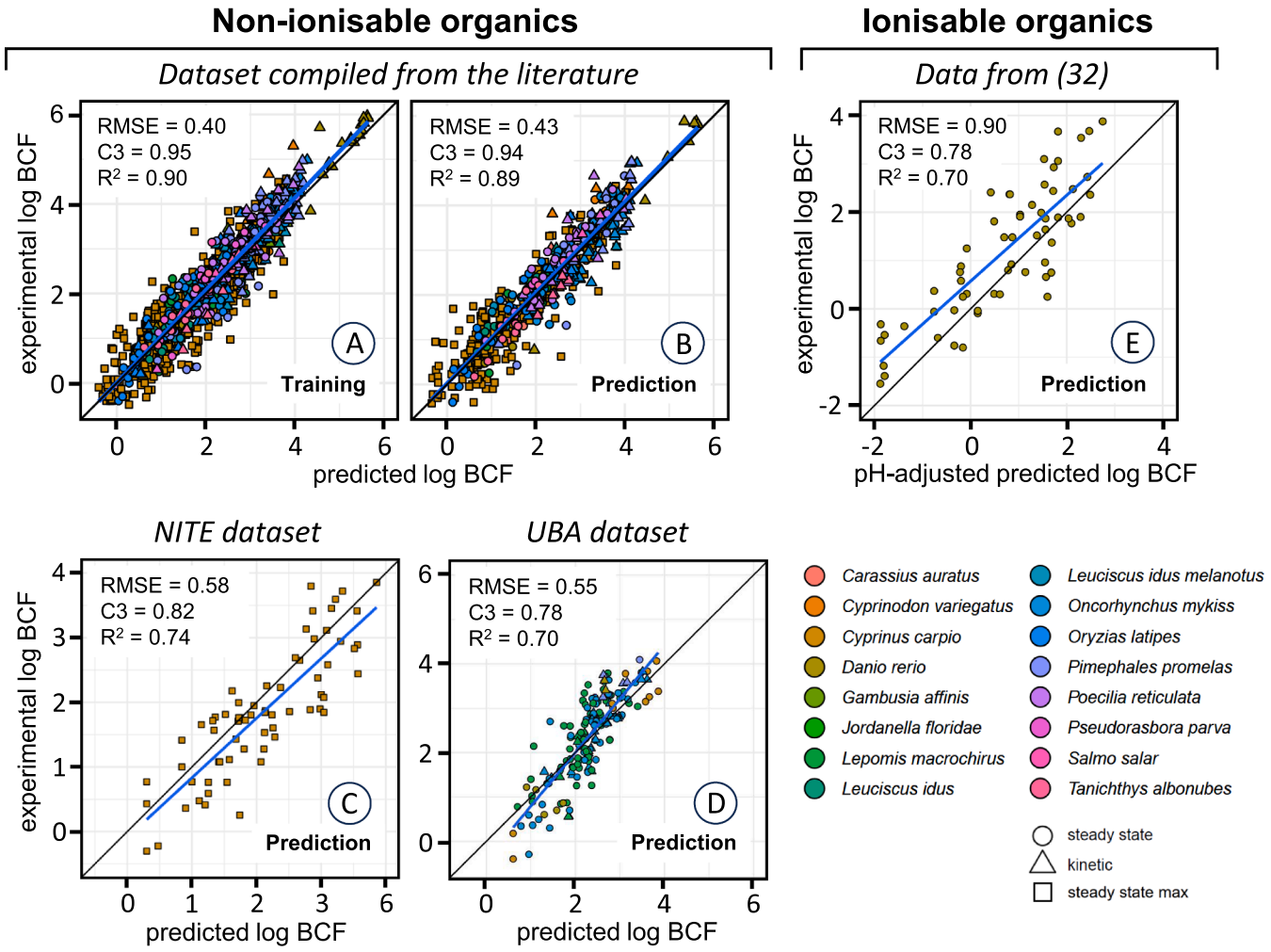



**Fig. 4.** Validation of AI-generated predictions for log BCF (x-axis) by experimentally collected log BCF (y-axis) for 16 fish species and more than a thousand chemicals under different exposure conditions. (A) AI-training subset with data for largely non-ionisable substances (n = 3032) from a literature-based dataset compiled for the present study. (B) Predictions for the remaining BCF data from the literature-compiled database not used for AI training (data for non-ionisable substances, n = 760). (C) Predictions for an additional publicly accessible BCF dataset from the Japanese National Institute of Technology and Evaluation (NITE) that originated from the Chemicals Evaluation and Research Institute (CERI), Japan, for *Cyprinus carpio* (data for largely non-ionisable substances, collected after several weeks of exposure presumably at steady state, n = 65). (D) Predictions for confidential BCF data submitted to the European Chemicals Agency (ECHA), the European Food Safety Authority (EFSA) and the European Medicines Agency (EMA) and handled by the German Environmental Agency (UBA) for authorization purposes (non-ionisable substances, n = 165). (E) Predictions for BCF data on 18 ionisable substances collected at 3–4 different pH levels for *Danio rerio* embryos, using AI-generated predictions and ΔlogD corrections [24] for pH-adjusted data calculation (n = 60). Linear regressions (blue lines). RMSE: root mean square error. C3: concordance correlation coefficient for agreement on a continuous measure and a measure of how well the data set fits the 1:1 line (black bisectors) [41, 42]. $R^2$: Bravais-Pearson variance. In all cases correlations were highly significant at $P < 0.0001$. Circles represent data presumably collected at steady state, triangles represent data calculated during the kinetics of progressive accumulation, squares represent maximum values presumably collected during steady state conditions (CERI). The different colours of the symbols refer to the tested species. In all cases, the model was able to predict unknown test data with an accuracy of 70–89 %. For chemicals with a BCF > 5000 (logBCF > 3.7), *D. rerio* was the species for which the most accurate predictions were possible, probably because the training of the AI at such high BCFs was largely based on experimental data obtained from zebrafish (see also supplementary Figs. S17-S19).

[53], and to propose a lower EU EQS for diclofenac [27], based on realistic worst-case conditions for ambient pH, as compared to actual test conditions. Using BCFpro we could also demonstrate that fish species sensitivity to bioaccumulation varies significantly depending on the specific substance and test conditions, rather than there being a universally most sensitive species, which has crucial implications for biodiversity conservation efforts. Given that our model successfully incorporates the most critical parameters other factors that were only available in a small fraction of studies (e.g. metabolization and excretion capacity of the fish, fat content, growth, freely dissolved concentration and total organic carbon (TOC) in the test system), were not taken into account in our AI modelling approach and are discussed in the SI.

### 3.4. How to meet regulatory demands?

From a regulatory perspective, bioaccumulative substances need to be identified at an early stage during environmental risk assessment prior to marketing to avoid long-term effects on wildlife, humans and ecosystems. Furthermore, emissions of such compounds need to be minimized as much as possible, as their long term-effects are still unpredictable [69]. Following this demand, the assessment of the bioaccumulation potential of chemicals has become a crucial component in the assessment of persistent (P), bioaccumulative (B) and toxic (T) chemicals. These PBT assessments are mandatory under various European legislations, such for industrial chemicals under REACH [70], for

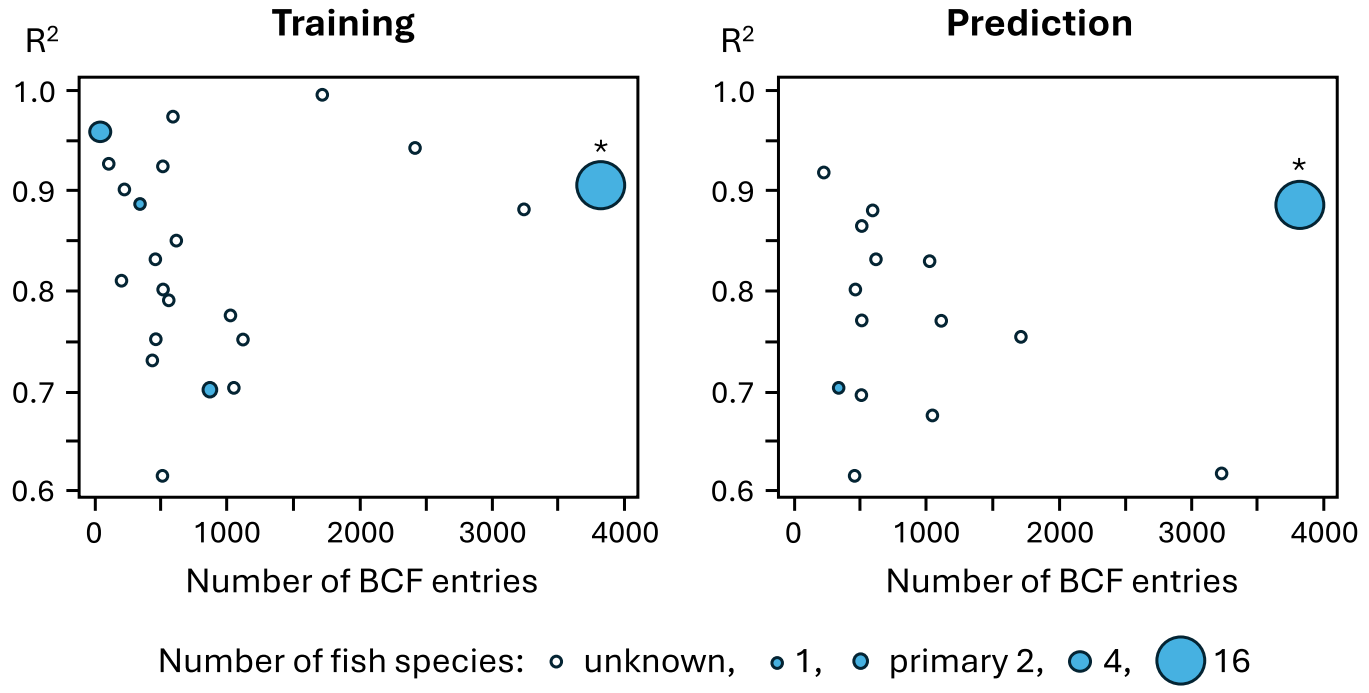



**Fig. 5.** Properties of existing models for the *in silico* generation of BCF data. The number of BCF entries used in the studies is shown against the coefficient of determination ($R^2$) obtained when modelling the training (left) and test (right, prediction) datasets. The size of the dots symbolises the number of fish species for which predictions are possible. Data taken from the following studies (in order of increasing number of BCF entries): Ivanciuc et al. (2006), Saçan et al. (2004), Lu et al. (2000), Toropova et al. (2013), Miller et al. (2019), Dimitrov et al. (2002), Piir et al. (2010), Zhao et al. (2008), Kobayashi & Yoshida (2021), Pramanik & Roy (2014), Weisbrod et al. (2007), Pore et al. (2024), Zang et al. (2017), Mansouri et al. (2018), Arnot & Quinn (2015), Toropov et al. (2011), Zhao et al. (2022), Lunghini et al. (2019), Zhu et al. (2025), Chandra et al. (2025), Xu et al. (2023) [10,14,21–23,30,54–68]. *: own study which is furthermore unique in that it incorporates several different BCF values for the same chemical, as found in the literature, into modelling on a larger scale. Details in the supplementary Table S2.

biocides [71], plant protection products [72], pharmaceuticals [73], more recently also under the regulation of Classification, Packaging and Labeling (CLP) [74], but also in the USA [75], Canada [76] and under international legislation such as the Stockholm Convention on Persistent Organic Pollutants (POPs) [77].

Experimentally generating reliable regulatory BCF values for the vast array of current anthropogenic substances poses an insurmountable challenge and will only continue to worsen in the future. A growing population will drive demand for chemicals and chemical-intensive products. Since 1950, chemical production has increased 50-fold. It is expected to again triple by 2050 [78], with more than 350,000 chemicals currently in use worldwide [79] and 100,000 in Europe [78, 80]. In addition, several thousands of largely unidentified (bio-)transformation products with unknown hazard profiles are formed in environmental media [81, 82], which should also be considered in a circular economy if a sustainable world is to be achieved.

The BCFpro algorithm stands out by allowing BCF values to be estimated with a high degree of confidence for 16 different fish species while at the same time quantifying the extent to which specific test conditions in the experimental test design explain the BCF value. Furthermore, BCFpro‘s ability to simultaneously calculate BCF values for numerous combinations of exposure conditions allows for the determination of a theoretical maximum BCF that can be achieved experimentally for a given organic chemical, and, consequently, the derivation of BCF90 and BCF50 values. Setting a percentage level (90 % for BCF90) as a guideline for regulatory applications may seem arbitrary at first. However, numerous examples exist in ecotoxicology and other fields where practical percentage values have been set that are intended to provide protection while tolerating a certain residual risk. These values are both scientifically and socially accepted. Examples include HC5 (5 %) in species sensitivity distribution analysis, SpO2 (90 % oxygen saturation in human blood for hypoxia), HbA1C (6.5 % for diabetes mellitus diagnosis) and 60 % of gross income product as the threshold for risky national debt. In this respect, it seems both practical and plausible to tolerate 10 % of the possible maximum value as residual risk when using a modelled BCF in regulation, and to set the 90 % level as the relevant metric.

So, in our opinion, the BCF90 can be regarded as a proxy for worst-case conditions that should be considered in regulatory frameworks. Encouraged by the exceptionally high success rate of BCFpro in predicting BCFs for unknown chemicals (see Fig. 4), we applied the generation of BCF90 and BCF50 values to chemicals submitted for registration with the EU, in order to verify the consistency of our approach with the risk assessment carried out by the authorities. However, an analysis of 165 BCF values submitted to EU authorities (ECHA, EFSA, EMA) for the purpose of chemical regulation revealed that 26.7 % of submitted values were more than 10 times lower than their corresponding BCF90 values (Fig. 7).

Almost a fifth of the submitted values (19.4 %) even fell short of the median metric, BCF50, by a factor of 10 or more. For chemicals classified as bioaccumulative, i.e. with an experimental BCF $>$ 1000 (USA) or $>$ 2000 (EU), BCFpro correctly identified 89.4 % (for BCF $>$1000) and 84.6 % (for BCF $>$2000), respectively, by calculating the BCF90. This sensitivity of up to almost 90 % shows the high reliability of the BCF90 as a worst-case metric. Alarmingly, however, 46.9 % and 62.7 % of the chemicals for which BCF90 values $>$ 1000 and $>$ 2000 are predicted by BCFpro, respectively, were reported to EU authorities to have BCF values below 1000 and 2000, so that these chemicals were not classified according to their true accumulation potential (Fig. 7).

### *3.5. Considerations regarding the limitations and potential improvements to BCFpro*

Finally, the limitations of the modelling primarily concern the raw data, and these are discussed in detail in the SI. For example, lipid normalisation is essential for producing consistent, comparable and

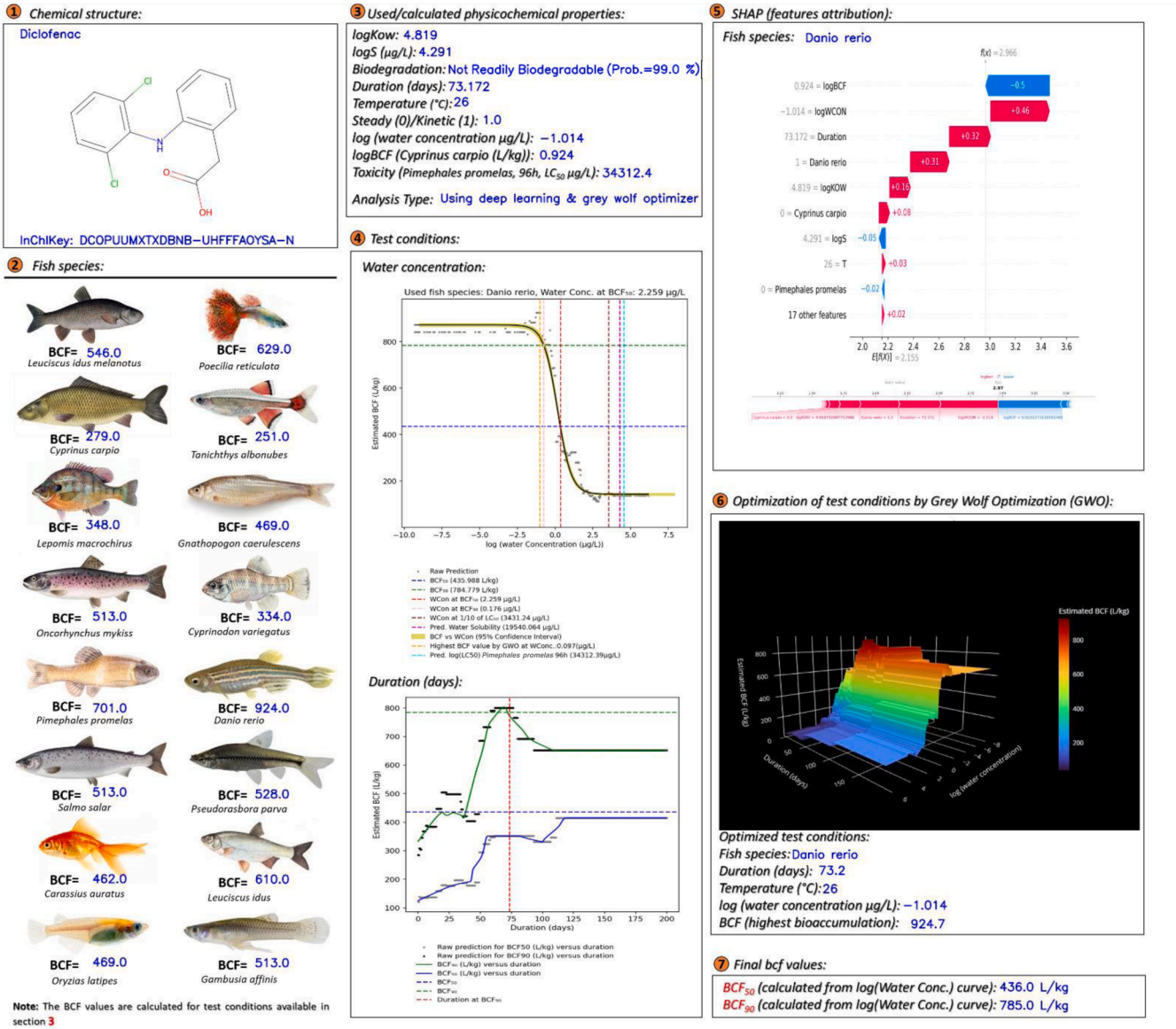



**Fig. 6.** Screenshot of BCFpro with visualized data - example for the anti-inflammatory drug diclofenac. (1): Chemical structure and IUPAC Standard International Chemical Identifier (InChI) Key. (2): Habitus of 16 test species and species-specifically predicted BCFs for those conditions that have been calculated as optimal for a selected species (here *Danio rerio*). (3): Test conditions-specific parameters used to calculate the resulting BCF for the 16 species (panel 2) under the optimal test conditions for *Danio rer*io in this example. (4): Predicted BCF (y-axis) vs. log water concentration (x-axis, top) and vs. exposure duration (x-axis, bottom). Top: The graph on water concentration vs BCF includes the estimated BCF90 (green dashed line) and BCF50 (blue dashed line), their corresponding water concentrations, the predicted maximum water solubility, the $LC_{50}$ and 0.1 x $LC_{50}$, and the maximum expected BCF. It also includes generated BCF predictions for specific water concentrations (raw data, gray dots) as well as the non-linear fit curve (black line) and its 95 % confidence interval (yellow-green area). Bottom: Calculation of the time course of predicted values representing 90 % and 50 % of the maximum BCF at each time point (black and gray dots) and the corresponding fitted curves (green and blue, respectively). Predicted BCF 90 (green dashed line) and BCF 50 (blue dashed line) at respective water concentrations are shown for comparison. The red dashed line indicates the exposure time predicted to reach the BCF90. (5): Shapley Additive Explanations (SHAPs) values explaining the effect of each variable for the predictions of BCFs. Since the initial DL model was trained to predict BCF data for *Cyprinus carpio* only, XGBoost applies various corrections to the initially calculated values for optimized prediction, considering water concentration, exposure duration, different fish species, etc. These corrections in positive and negative directions are visualized here. (6): Three-dimensional plot of test condition-specific BCFs as a function of water concentration and exposure time for the fish species for which the maximum BCF was determined (here: *Danio rerio*). Indication of the parameters for which this maximum BCF can be achieved experimentally. (7): Predicted BCF90 and BCF50 values for diclofenac (selected chemical) and *Danio rerio* (species with maximum BCF).

meaningful bioconcentration data in fish studies [83], particularly with regard to lipophilic substances. This accounts for natural variability in fish lipid content, which would otherwise make it difficult to interpret bioaccumulation results. According to OECD TG 305, the BCF must be corrected for a lipid content of 5 % (for both steady-state and kinetic BCF) throughout the study period, in order to minimise variation in BCF values [8]. However, the lipid content of the fish was not reported for the majority (>80 %) of the available experimental data. In the publicly available Japanese NITE dataset, the lipid content of the fish was not reported at all. Therefore, the effect of fish lipid content could not be included in the BCFpro modelling and should be addressed when more data becomes available to further increase the reliability of BCF predictions. Furthermore, other biological and technical parameters potentially affect BCF values in experimental studies (e.g. growth correction and species capacity for the metabolism and excretion of chemicals; see SI). They should be considered and implemented as

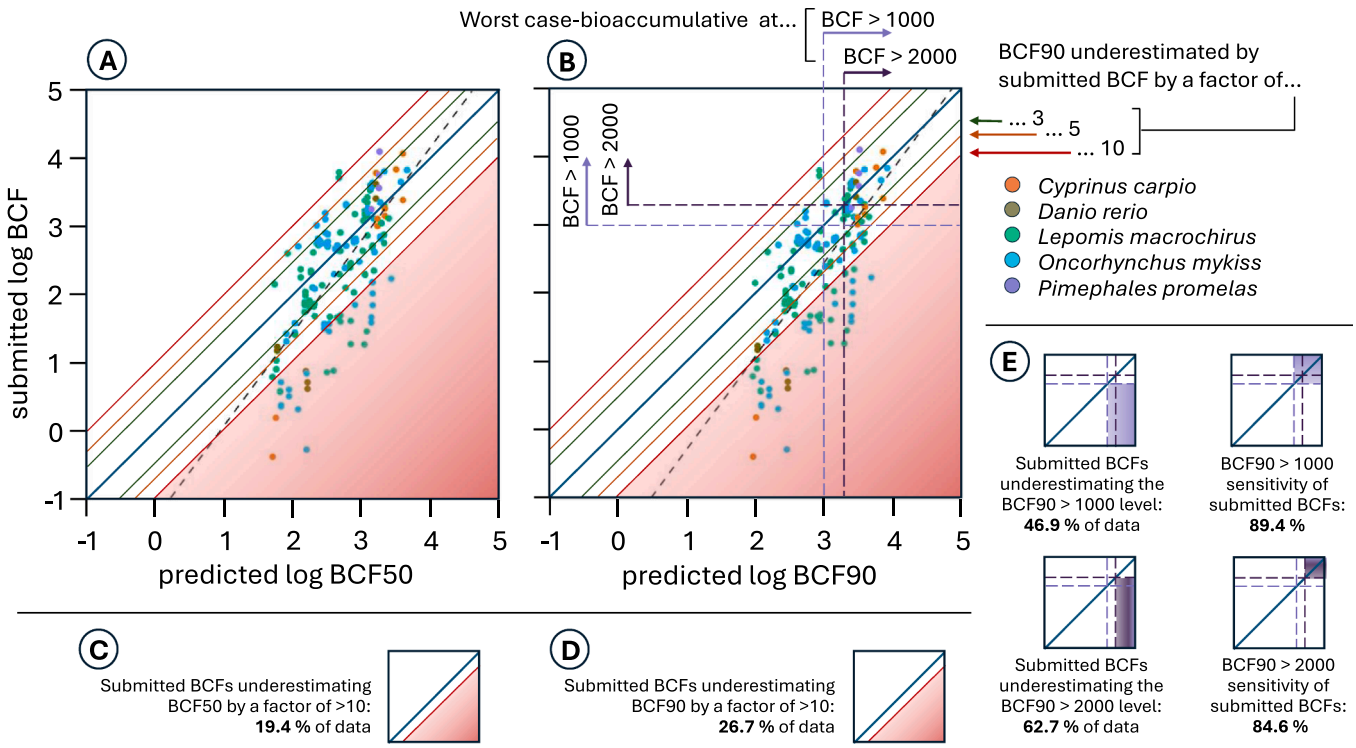



**Fig. 7.** Comparison of BCF values for organic chemicals and fish from randomly selected studies submitted to the EU for chemical regulation with BCF50 and BCF90 values calculated for these substances using BCFpro (n = 165, chemical identity confidential). (A): BCFpro predicted log BCF50 *vs.* submitted BCF. Linear regression line in black, dashed. Blue line: bisector (1:1 ratio). Over- and underestimation of BCF50 by the submitted studies by factors of 3, 5 and 10 are symbolized by lines parallel to the bisector in green, orange, red, respectively. The different colors of the symbols refer to the tested species. (B): Same plot as in (A) but comparing BCFpro-derived predicted log BCF90 *vs.* submitted BCF. The vertical and horizontal light and dark violet dashed lines symbolize the limits above which a substance is considered to be bioaccumulative (1000 for USA, Japan, 2000 for Australia, EU). (C): Proportion of submitted BCF values that underestimate the BCF50 by more than a factor of 10 (light red part of graph (A), 32 of 165 data points, corresponding to 19.4 %). (D): Percentage of submitted BCF values that underestimate the BCF90 (worst-case bioaccumulation factor) by more than a factor of 10 (light red colored part of graph (B), 44 out of 165 data points, 26.7 %). (E): Top left: ratio of chemicals with submitted BCF data < 1000 to those with BCF90 > 1000 (i.e. 1 - predictivity). BCFpro determines a BCF90 > 1000 for 81 substances, but for 38 of them a BCF < 1000 was submitted (underestimation by 46.9 % of these data). (E) Lower left: the same proportion of submitted BCFs < 2000 for those substances for which BCFpro determined a BCF90 > 2000. This corresponds to 37 out of 59 data points underestimating the BCF90 (i.e. 62.7 %). (E) Top right: sensitivity (i.e. the proportion of actual positives that are correctly identified) of BCF90 > 1000 for submitted BCF data > 1000. For 42 out of 47 chemicals, both metrics are above 1000 (89.4 %). (E) Lower right: sensitivity of BCF90 > 2000 for submitted BCF data > 2000. Twenty-two out of 26 chemicals have both metrics above 2000 (84.6 %). The areas where the corresponding data points in plot (B) are located are indicated by the violet-colored sectors in the small graphs in (E).

additional factors in AI models such as BCFpro in future to increase reliability. Therefore, we strongly recommend that future bioconcentration studies describe the experimental design and test conditions in detail and perform tests according to realistic worst-case conditions (justified by the BCFpro model), in order to support sound interpretation and robust safety assessments. This will help continuously update the available 'gold standard' BCF database, improve the training data and increase the accuracy of future bioaccumulation models, enabling comparison with quality-assured benchmark studies. Further research is needed to improve explicability and regulatory acceptance. Transparency, comparability and reproducibility beyond individual studies must also be considered when developing future models.

## 4. Conclusions

We can identify three areas in which the novel, standardised metrics and the BCFpro application can be put into practice:

(1) Support of regulatory authorities and risk assessment: (a) Review of dossiers submitted to authorities for authorisation or registration of chemicals that have chosen rather arbitrary test conditions to determine one BCF among many possible ones. As our developed tool is a fast and very cost-effective alternative to the OECD TG 305 bioaccumulation test, the BCF90 can be quickly and easily calculated and used as a realistic worst-case surrogate, especially regarding environmentally relevant concentrations. Furthermore, we suggest (in analogy to the $LC_{50}$) that the mean value in a continuum of possible bioaccumulation factors, i.e. the BCF50, should be considered as a minimum requirement in this context. (b) Defining specifications for fish species and specific test conditions such as chemical concentration and exposure time for the experimental determination of meaningful bioconcentration factors. This ensures more standardised and accurate bioaccumulation assessments in the future. The results of this study have been shared with relevant regulatory authorities, most notably with the European Chemicals Agency, for further discussion and have the potential to stipulate further activity towards future implementation in chemical risk assessment guidelines and legislation.

(2) Support for chemical manufacturers and societal and ethical significance: numerically unlimited screening of ecotoxicologically uncharacterised chemicals and their metabolites prior to registration for their bioaccumulation potential aimed at identifying them *a priori* as potential B or vB substances, which can significantly contribute to a global reduction of animal experiments. Thus, our approach has great potential as an *in silico* alternative to fish testing. It combines cost savings with social and ethical benefits.

(3) Implications for environmental monitoring programs: Understanding a substance's or organism's high bioaccumulation factor enables us to focus monitoring efforts on those that pose the greatest long-term risks to ecosystems and health. Substances with high BCFs should be prioritised for monitoring, as they can cause ecological damage in the long term even at low concentrations in the environment. BCF values can inform prioritisation lists, enabling the limited resources available for monitoring programmes to be used more effectively to maximise environmental benefit. By reducing the uncertainty in bioaccumulation estimates, AI-based models such as BCFpro will ultimately strengthen the ecological realism of risk assessments. This is achieved by predicting true body burdens, which can then be linked to adverse effects such as reproduction, survival and population viability. This will ultimately help us to understand the impact on biodiversity and its implications for risk management under regulatory frameworks such as REACH and the Water Framework Directive.

## Environmental Implication

Determining bioaccumulation factors (BCFs) is an essential part of the chemical authorization process and provides information on their potential to accumulate in environmental compartments. Our work shows that a chemical's BCF is highly dependent on its concentration in the external medium, which leaves great scope for reporting BCFs to authorities. However, a standardized BCF value that reflects worst-case scenarios would reliably classify bioaccumulating substances. In our study, we introduce such a novel metric, called BCF90, and present an AI-generated methodology that can determine this metric for organic environmental chemicals with up to 90 % accuracy *in silico*.

## CRediT authorship contribution statement

**Reza Aalizadeh:** Writing – original draft, Visualization, Software, Methodology, Investigation. **Heinz-R. Köhler:** Writing – review & editing, Writing – original draft, Visualization, Supervision, Project administration, Investigation, Funding acquisition, Conceptualization. **Peter C. von der Ohe:** Writing – review & editing, Writing – original draft, Supervision, Methodology, Investigation, Conceptualization. **Rita Triebskorn:** Writing – review & editing, Supervision, Funding acquisition. **Nikolaos S. Thomaidis:** Supervision, Funding acquisition. **Ines Prutz:** Writing – review & editing, Investigation. **Katharina Peschke:** Visualization, Project administration, Investigation, Funding acquisition, Data curation. **Thomas Gräff:** Visualization, Investigation. **Gabriele Treu:** Writing – review & editing, Writing – original draft, Investigation.

## Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Acknowledgments

This work was supported by the German Environmental Agency [grant numbers FKZ 3718 67 415 0 (pHION) and 173534 - 97 292/0001 (BCF-Screen)] and by the TerraChem project, funded by the European Union's Horizon Europe research and innovation program under grant agreement No. 101135483. We thank Caren Rauert, Doris Hirmann and Yuki Sakuratani for discussions on regulatory issues, Eleni I. Panagopoulou and Dimitrios E. Damalas for analysis of ionizable substances in water and fish embryos, Stefanie Krais, Dominik Nerlich and Alexander von der Ohe for supporting computational work, Manuela Kühne for linguistic improvement of the manuscript and three anonymous reviewers for their comments on a previous version of this paper. Reza Aalizadeh declares that this work was conducted while the author was affiliated with the Laboratory of Analytical Chemistry, Department of Chemistry, National and Kapodistrian University of Athens, Panepistimiopolis Zografou, 15771 Athens, Greece. His current address is Department of Environmental Health Sciences, Yale School of Public Health, Yale University, New Haven, Connecticut 06510, U.S.A. The conclusions expressed in this paper represent the expert judgement of the authors, but not necessarily the opinion of their affiliation.

## Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.jhazmat.2025.140648.

## Data Availability

Data are available from the corresponding authors upon request. BCFpro software is distributed free of charge at: https://www.parcopedia.eu/BCFpro . Apart from this, all data are available in the main text or the supplementary materials.

## References

[1] Cade, T.J., Lincer, J.L., White, C.M., Roseneau, D.G., Swartz, L.G., 1971. DDE residues and eggshell changes in Alaskan falcons and hawks. Science 172, 955–957. https://doi.org/10.1126/science.172.3986.955.

[2] de Wit, C.A., Bossi, R., Dietz, R., Dreyer, A., Faxneld, S., Garbus, S.E., Hellström, P., Koschorreck, J., Lohmann, N., Roos, A., Sellström, U., Sonne, C., Treu, G., Vorkamp, K., Yuan, B., Eulaers, I., 2020. Organohalogen compounds of emerging concern in Baltic Sea biota: Levels, biomagnification potential and comparisons with legacy contaminants. Environ Int 144, 106037. https://doi.org/10.1016/j.envint.2020.106037.

[3] Desforges, J.-P., Hall, A., McConnell, B., Eulaers, I., Sonne, C., Rosing Asvid, A., Vikingson, G., Samarra, F., Jepson, P., Dietz, R., 2018. Projecting the global fate of killer whales from PCB pollution. Science.

[4] Sonne, C., Dietz, R., Jenssen, B.M., Lam, S.S., Letcher, R.J., 2021. Emerging contaminants and biological effects in Arctic wildlife. Trends Ecol Evol 36, 421–429. https://doi.org/10.1016/j.tree.2021.01.007.

[5] Treu, G., Slobodnik, J., Alygizakis, N., Badry, A., Bunke, D., Cincinelli, A., Claßen, D., Dekker, R.W.R.J., Göckener, B., Gkotsis, G., Hanke, G., Duke, G., Jartun, M., Movalli, P., Nika, M.-C., Rüdel, H., Tarazona, J.V., Thomaidis, N.S., Tornero, V., Vorkamp, K., Walker, L.A., Koschorreck, J., Dulio, V., 2022. Using environmental monitoring data from apex predators for chemicals management: towards better use of monitoring data from apex predators in support of prioritisation and risk assessment of chemicals in Europe. Environ Sci Eur 34, 82. https://doi.org/10.1186/s12302-022-00665-5.

[6] Fu, W., Franco, A., Trapp, S., 2009. Methods for estimating the bioconcentration factor of ionizable organic chemicals. Environ Toxicol Chem 28, 1372–1379. https://doi.org/10.1897/08-233.1.

[7] Wassenaar, P.N.H., Verbruggen, E.M.J., Cieraad, E., Peijnenburg, W.J.G.M., Vijver, M.G., 2020. Variability in fish bioconcentration factors: Influences of study design and consequences for regulation. Chemosphere 239, 124731. https://doi.org/10.1016/j.chemosphere.2019.124731.

[8] OECD, 2017. Guidance Document on Aspects of OECD TG 305 on Fish Bioaccumulation. ⟨https://www.oecd-ilibrary.org/content/publication/c25ad843-en⟩.

[9] Gimeno, S., Allan, D., Kai, P., Remuzat, P., Collard, M., 2024. Are current regulatory log Kow cut-off values fit-for-purpose as a screening tool for bioaccumulation potential in aquatic organisms? Regul Toxicol Pharmacol 147. https://doi.org/10.1016/j.yrtph.2023.105556.

[10] Weisbrod, A.V., Burkhard, L.P., Arnot, J., Mekenyan, O., Howard, P.H., Russom, C., Boethling, R., Sakuratani, Y., Traas, T., Bridges, T., Lutz, C., Bonnell, M., Woodburn, K., Parkerton, T., 2007. Workgroup report: review of fish bioaccumulation databases used to identify persistent, bioaccumlative, toxic substances. Environ Health Perspect 115, 255–261. https://doi.org/10.1289/ehp.9424.

[11] Zhao, L., Montanari, F., Heberle, H., Schmidt, S., 2022. Modeling bioconcentration factors in fish with explainable deep learning. Artif Intell Life Sci 2, 100047. https://doi.org/10.1016/j.ailsci.2022.100047.

[12] OECD, 2023. QSAR Assessment Framework: Guidance for the regulatory assessment of (Quantitative) Structure – Activity Relationship models, predictions, and results based on multiple predictions. In: Series on Testing and Assessment No, 386. ENV/CBC/MONO(2023), p. 32.

[13] Gasser, L., Schür, C., Perez-Cruz, F., Schirmer, K., Baity-Jesi, M., 2024. Machine learning-based prediction of fish acute mortality: implementation, interpretation, and regulatory relevance. Environ Sci Adv 3, 1124–1138. https://doi.org/10.1039/D4VA00072B.

[14] Pore, S., Pelloux, A., Chatterjee, M., Banerjee, A., Roy, K., 2024. Machine learning-based q-RASAR predictions of the bioconcentration factor of organic molecules estimated following the organisation for economic co-operation and development guideline 305. J Hazard Mater 479, 135725. https://doi.org/10.1016/j.jhazmat.2024.135725.
[15] von Borries, K., Holmquist, H., Kosnik, M., Beckwith, K.V., Jolliet, O., Goodman, J. M., Fantke, P., 2023. Potential for machine learning to address data gaps in human toxicity and ecotoxicity characterization. Environ Sci Technol 57, 18259–18270. https://doi.org/10.1021/acs.est.3c05300.
[16] Evangelista, D., Nelson, E., Skyner, R., Tehan, B., Bernetti, M., Roberti, M., Bolognesi, M.L., Bottegoni, G., 2025. Application of deep learning to predict the persistence, bioaccumulation, and toxicity of pharmaceuticals. J Chem Inf Model 65, 3248–3261. ⟨https://pubs.acs.org/doi/10.1021/acs.jcim.4c02293⟩.
[17] Ball, P., 2023. Is AI leading to a reproducibility crisis in science? Nature 624, 22–25. https://doi.org/10.1038/d41586-023-03817-6.
[18] Kretschmer, F., Seipp, J., Ludwig, M., Klau, G.W., Böcker, S., 2025. Coverage bias in small molecule machine learning. Nat Commun 16, 554. https://doi.org/10.1038/s41467-024-55462-w.
[19] Grisoni, F., Consonni, V., Villa, S., Vighi, M., Todeschini, R., 2015. QSAR models for bioconcentration: is the increase in the complexity justified by more accurate predictions? Chemosphere 127, 171–179. https://doi.org/10.1016/j.chemosphere.2015.01.047.
[20] Zhu, M., Xiao, Z., Zhang, T., Lu, G., 2025. Construction of interpretable ensemble learning models for predicting bioaccumulation parameters of organic chemicals in fish. J Hazard Mater 482, 136606. https://doi.org/10.1016/j.jhazmat.2024.136606.
[21] Ivanciuc, T., Ivanciuc, O., Klein, D.J., 2006. Modeling the bioconcentration factors and bioaccumulation factors of polychlorinated biphenyls with posetic quantitative super-structure/activity relationships (QSSAR). Mol Divers 10, 133–145. https://doi.org/10.1007/s11030-005-9003-3.
[22] Toropov, A.A., Toropova, A.P., Lombardo, A., Roncaglioni, A., Benfenati, E., Gini, G., 2011. CORAL: Building up the model for bioconcentration factor and defining it's applicability domain. Eur J Med Chem 46, 1400–1403. https://doi.org/10.1016/j.ejmech.2011.01.018.
[23] Lunghini, F., Marcou, G., Azam, P., Patoux, R., Enrici, M.H., Bonachera, F., Horvath, D., Varnek, A., 2019. QSPR models for bioconcentration factor (BCF): are they able to predict data of industrial interest? SAR QSAR Environ Res 30, 507–524. https://doi.org/10.1080/1062936X.2019.1626278.
[24] Xu, J.-Y., Wang, K., Men, S.-H., Yang, Y., Zhou, Q., Yan, Z.-G., 2023. QSAR-QSIIR-based prediction of bioconcentration factor using machine learning and preliminary application. Environ Int 177, 108003. https://doi.org/10.1016/j.envint.2023.108003.
[25] Arnot, J.A., Gobas, F.A.P.C., 2006. A review of bioconcentration factor (BCF) and bioaccumulation factor (BAF) assessments for organic chemicals in aquatic organisms. Environ Rev 14, 257–297, 10.1139/A06-005.
[26] Rodriguez-Sanchez, N., Cronin, M.T.D., Lillicrap, A., Madden, J.C., Piechota, P., Tollefsen, K.E., 2014. Development of a list of reference chemicals for evaluating alternative methods to in vivo fish bioaccumulation tests. Environ Toxicol Chem 33, 2740–2752. /10.1002/etc.2734.
[27] Kroll, A., von der Ohe, P.C., Köhler, H.-R., Sellier, O., Junghans, M., 2024. Aquatic thresholds for ionisable substances, such as diclofenac, should consider pH-specific differences in uptake and toxicity. Sci Total Environ 908, 168222. https://doi.org/10.1016/j.scitotenv.2023.168222.
[28] Oelkers, K., 2021. Transparency in product authorisation – the handling of commercially confidential information in environmental risk assessments of industrial chemicals, biocidal products, plant protection products and pharmaceuticals. Sustain Chem Pharm 20, 100399. https://doi.org/10.1016/j.scp.2021.100399.
[29] Huang, K., Zhang, H., 2022. Classification and regression machine learning models for predicting aerobic ready and inherent biodegradation of organic chemicals in water. Environ Sci Technol 56, 12755–12764. https://doi.org/10.1021/acs.est.2c01764.
[30] Mansouri, K., Grulke, C.M., Judson, R.S., Williams, A.J., 2018. OPERA models for predicting physicochemical properties and environmental fate endpoints. J Chemin 10. https://doi.org/10.1186/s13321-018-0263-1.
[31] Mirjalili, S., Mirjalili, S.M., Lewis, A., 2014. Grey wolf optimizer. Adv Eng Softw 69, 46–61. https://doi.org/10.1016/j.advengsoft.2013.12.007.
[32] Köhler, H.-R., Gräff, T., Schweizer, M., Blumhardt, J., Burkhardt, J., Ehmann, L., Hebel, J., Heid, C., Kundy, L., Kuttler, J., Malusova, M., Moroff, F.-M., Schlösinger, A.-F., Schulze-Berge, P., Panagopoulou, E.I., Damalas, D.E., Thomaidis, N.S., Triebskorn, R., Maletzki, D., Kühnen, U., von der Ohe, P.C., 2023. LogD-based modelling and ΔlogD as a proxy for pH-dependent action of ionizable chemicals reveal the relevance of both neutral and ionic species for fish embryotoxicity and possess great potential for practical application in the regulation of chemicals. Water Res 235, 119864. https://doi.org/10.1016/j.watres.2023.119864.
[33] Isnard, P., Lambert, S., 1988. Estimating bioconcentration factors from octanol-water partition coefficient and aqueous solubility. Chemosphere 17, 21–34. https://doi.org/10.1016/0045-6535(88)90040-9.
[34] Mackay, D., Arnot, J.A., Gobas, F.A.P.C., Powell, D.E., 2013. Mathematical relationships between metrics of chemical bioaccumulation in fish. Environ Toxicol Chem 32, 1459–1466. https://doi.org/10.1002/etc.2205.
[35] Mackay, D., 1982. Correlation of bioconcentration factors. Environ Sci Technol 16, 274–278. https://doi.org/10.1021/es00099a008.
[36] Chiou, C.T., Freed, V.H., Schmedding, D.W., Kohnert, R.L., 1977. Partition coefficient and bioaccumulation of selected organic chemicals. Environ Sci Technol 11, 475–478. https://doi.org/10.1021/es60128a001.
[37] OECD, 2012. Test No. 305: Bioaccumulation in Fish: Aqueous and Dietary Exposure, OECD Guidelines for the Testing of Chemicals, Section 3. OECD Publishing, Paris. https://doi.org/10.1787/9789264185296-en.
[38] Creton, S., Weltje, L., Hobson, H., Wheeler, J.R., 2013. Reducing the number of fish in bioconcentration studies for plant protection products by reducing the number of test concentrations. Chemosphere 90, 1300–1304. https://doi.org/10.1016/j.chemosphere.2012.09.029.
[39] Burden, N., Creton, S., Weltje, L., Maynard, S.K., Wheeler, J.R., 2014. Reducing the number of fish in bioconcentration studies with general chemicals by reducing the number of test concentrations. Regul Toxicol Pharmacol 70, 442–445. https://doi.org/10.1016/j.yrtph.2014.08.008.
[40] Sherman, W.R., Keenan, R.E., Gunster, D.G., 1992. Reevaluation of dioxin bioconcentration and bioaccumulation factors for regulatory purposes. J Toxicol Environ Health 37, 211–229. https://doi.org/10.1080/15287399209531666.
[41] Fick, J., Lindberg, R.H., Parkkonen, J., Arvidsson, B., Tysklind, M., Larsson, D.G.J., 2010. Therapeutic levels of levonorgestrel detected in blood plasma of fish: results from screening rainbow trout exposed to treated sewage effluents. Environ Sci Technol 44, 2661–2666. https://doi.org/10.1021/es903440m.
[42] Mehinto, A.C., Hill, E.M., Tyler, C.R., 2010. Uptake and biological effects of environmentally relevant concentrations of the nonsteroidal anti-inflammatory pharmaceutical diclofenac in rainbow trout (Oncorhynchus mykiss). Environ Sci Technol 44, 2176–2182. https://doi.org/10.1021/es903702m.
[43] Nichols, J.W., Du, B., Berninger, J.P., Connors, K.A., Chambliss, C.K., Erickson, R. J., Hoffman, A.D., Brooks, B.W., 2015. Observed and modeled effects of pH on bioconcentration of diphenhydramine, a weakly basic pharmaceutical, in fathead minnows. Environ Toxicol Chem 34, 1425–1435. https://doi.org/10.1002/etc.2948.
[44] Wu, M., Pan, C., Chen, Z., Jiang, L., Lei, P., Yang, M., 2017. Bioconcentration pattern and induced apoptosis of bisphenol A in zebrafish embryos at environmentally relevant concentrations. Environ Sci Pollut Res 24, 6611–6621. https://doi.org/10.1007/s11356-016-8351-0.
[45] Zhang, R., Wang, X., Zhang, X., Song, C., Letcher, R.J., Liu, C., 2018. Polychlorinated diphenylsulfides activate aryl hydrocarbon receptor 2 in Zebrafish embryos: potential mechanism of developmental toxicity. Environ Sci Technol 52, 4402–4412. https://doi.org/10.1021/acs.est.8b00366.
[46] Gobas, F.A.C.P., Morrison, H.A., 2000. Bioconcentration and biomagnification in the aquatic environment. In: Boethling, R.S., Mackay, D. (Eds.), Handbook of property estimation methods for chemicals. environmental and health sciences. Boca Raton, USA, pp. 189–231.
[47] Schwarzenbach, R.P., Gschwend, P.M., Imboden, D.M., 2003. Transformation Processes. Wiley Online Library.
[48] Ebert, A., Ackermann, J., Goss, K.-U., 2023. Mechanistic modeling of the bioconcentration of (super)hydrophobic compounds in Hyalella azteca. Environ Sci Pollut Res 30, 50257–50268. https://doi.org/10.1007/s11356-023-25827-7.
[49] Barron, M.G., 1990. Bioconcentration. Will water-borne organic chemicals accumulate in aquatic animals? Environ Sci Technol 24, 1612–1618. https://doi.org/10.1021/es00081a001.
[50] LeBlanc, G.A., 1995. Trophic-level differences in the bioconcentration of chemicals: implications in assessing environmental biomagnification. Environ Sci Technol 29, 154–160. https://doi.org/10.1021/es00001a020.
[51] Sijm, D.T.H.M., Pärt, P., Opperhuizen, A., 1993. The influence of temperature on the uptake rate constants of hydrophobic compounds determined by the isolated perfused gills of rainbow trout (Oncorhynchus mykiss). Aquat Toxicol 25, 1–14. https://doi.org/10.1016/0166-445X(93)90016-T.
[52] Schlechtriem, C., Fliedner, A., Schäfers, C., 2012. Determination of lipid content in fish samples from bioaccumulation studies: contributions to the revision of guideline OECD 305. Environ Sci Eur 24, 13. https://doi.org/10.1186/2190-4715-24-13.
[53] Scheer, 2022. Scientific Committee on Health, Environmental and Emerging Risks (SCHEER). Final opinion on draft environmental quality standards for priority substances under the water framework directive – ibuprofen, 5 December 2022.
[54] Saçan, M.T., Erdem, S.S., Özpınar, G.A., Balcıoglu, I.A., 2004. QSPR study on the bioconcentration factors of nonionic organic compounds in fish by characteristic root index and semiempirical molecular descriptors. J Chem Inf Comput Sci 44, 985–992. https://doi.org/10.1021/ci0342167.
[55] Lu, X., Tao, S., Hu, H., Dawson, R.W., 2000. Estimation of bioconcentration factors of nonionic organic compounds in fish by molecular connectivity indices and polarity correction factors. Chemosphere 41, 1675–1688. https://doi.org/10.1016/S0045-6535(00)00050-3.
[56] Toropova, A.P., Toropov, A.A., Martyanov, S.E., Benfenati, E., Gini, G., Leszczynska, D., Leszczynski, J., 2013. CORAL: monte carlo method as a tool for the prediction of the bioconcentration factor of industrial pollutants. Mol Inform 32, 145–154. https://doi.org/10.1002/minf.201200069.
[57] Miller, T.H., Gallidabino, M.D., MacRae, J.R., Owen, S.F., Bury, N.R., Barron, L.P., 2019. Prediction of bioconcentration factors in fish and invertebrates using machine learning. Sci Total Environ 648, 80–89. https://doi.org/10.1016/j.scitotenv.2018.08.122.
[58] Dimitrov, S.D., Dimitrova, N.C., Walker, J.D., Veith, G.D., Mekenyan, O.G., 2002. Predicting bioconcentration factors of highly hydrophobic chemicals. Effects of molecular size. Pure Appl Chem 74, 1823–1830. https://doi.org/10.1351/pac200274101823.
[59] Piir, G., Sild, S., Roncaglioni, A., Benfenati, E., Maran, U., 2010. QSAR model for the prediction of bio-concentration factor using aqueous solubility and descriptors

considering various electronic effects. SAR QSAR Environ Res 21, 711–729. https://doi.org/10.1080/1062936X.2010.528596.
[60] Zhao, C., Boriani, E., Chana, A., Roncaglioni, A., Benfenati, E., 2008. A new hybrid system of QSAR models for predicting bioconcentration factors (BCF). Chemosphere 73, 1701–1707. https://doi.org/10.1016/j.chemosphere.2008.09.033.
[61] Kobayashi, Y., Yoshida, K., 2021. Development of QSAR models for prediction of fish bioconcentration factors using physicochemical properties and molecular descriptors with machine learning algorithms. Ecol Inform 63, 101285. https://doi.org/10.1016/j.ecoinf.2021.101285.
[62] Pramanik, S., Roy, K., 2014. Modeling bioconcentration factor (BCF) using mechanistically interpretable descriptors computed from open source tool "PaDEL-Descriptor. Environ Sci Pollut Res 21, 2955–2965. https://doi.org/10.1007/s11356-013-2247-z.
[63] Zang, Q., Mansouri, K., Williams, A.J., Judson, R.S., Allen, D.G., Casey, W.M., Kleinstreuer, N.C., 2017. In silico prediction of physicochemical properties of environmental chemicals using molecular fingerprints and machine learning. J Chem Inf Model 57, 36–49. https://doi.org/10.1021/acs.jcim.6b00625.
[64] Arnot, J.A., Quinn, C.L., 2015. Development and evaluation of a database of dietary bioaccumulation test data for organic chemicals in fish. Environ Sci Technol 49, 4783–4796. https://doi.org/10.1021/es506251q.
[65] Zhao, L., Montanari, F., Heberle, H., Schmidt, S., 2022. Modeling bioconcentration factors in fish with explainable deep learning. Artif Intell Life Sci 2, 100047. https://doi.org/10.1016/j.ailsci.2022.100047.
[66] Zhu, M., Xiao, Z., Zhang, T., Lu, G., 2025. Construction of interpretable ensemble learning models for predicting bioaccumulation parameters of organic chemicals in fish. J Hazard Mater 482, 136606. https://doi.org/10.1016/j.jhazmat.2024.136606.
[67] Chandra, A., Banerjee, A., Roy, K., 2025. Prediction of bioconcentration factors (BCFs) and bioaccumulation factors (BAFs) for per- and polyfluoroalkyl substances (PFASs) using Read-Across and q-RASPR. Sci Total Environ 993, 180007. https://doi.org/10.1016/j.scitotenv.2025.180007.
[68] Xu, J.-Y., Wang, K., Men, S.-H., Yang, Y., Zhou, Q., Yan, Z.-G., 2023. QSAR-QSIIR-based prediction of bioconcentration factor using machine learning and preliminary application. Environ Int 177, 108003. https://doi.org/10.1016/j.envint.2023.108003.
[69] ECHA, 2023. Guidance on Information Requirements and Chemical Safety Assessment. Chapter R.11: PBT/vPvB assessment. Version 4.0.
[70] EC, 2006. Regulation (EC) No 1907/2006 of the European Parliament and of the Council of 18 December 2006 concerning the Registration, Evaluation, Authorisation and Restriction of Chemicals (REACH). J Eur Union L 136, 3–280.
[71] EC, 2012. Regulation (EU) No 528/2012 of the European Parliament and of the Council of 22 May 2012 concerning the making available on the market and use of biocidal products Text with EEA relevance.
[72] EC, 2009. Regulation (EC) No 1107/2009 of the European Parliament and of the Council of 21 October 2009 concerning the placing of plant protection products on the market and repealing Council Directives 79/117/EEC and 91/414/EEC.
[73] EC, 2004. Regulation (EC) No 726/2004 of the European Parliament and of the Council of 31 March 2004 laying down Community procedures for the authorisation and supervision of medicinal products for human and veterinary use and establishing a European Medicines Agency (Text with EEA relevance).
[74] EC, 2008. Regulation (EC) No 1272/2008 of the European Parliament and of the Council of 16 December 2008 on classification, labelling and packaging of substances and mixtures, amending and repealing Directives 67/548/EEC and 1999/45/EC, and amending Regulation (EC) No 1907/2006 (Text with EEA relevance).
[75] TSCA, 1976. The Toxic Control Act. Enacted by the 94th United States Congress October 11, 1976. Public Law 94-469, U.S.A.
[76] CEPA, 1999. Canadian Environmental Protection Act S.C. 1999, c. 33, Assented to 1999-09-14.
[77] United Nations. Persistent Organic Pollutants Review Committee (POPRC-7) of the Stockholm Convention on Persistent Organic Pollutants (POPs) 2011. Stockholm Convention on Persistent Organic Pollutants, signed on 22 May 2001 and effective from 17 May 2004.
[78] EEA, 2019. The European environment - state and outlook 2020. Knowledge for transition to a sustainable Europe. European Environment Agency (EEA). Luxembourg: Publications Office of the European Union 2019; Chapter 10 - Chemical pollution.
[79] Wang, Z., Walker, G.W., Muir, D.C.G., Nagatani-Yoshida, K., 2020. Toward a global understanding of chemical pollution: a first comprehensive analysis of national and regional chemical inventories. Environ Sci Technol 54, 2575–2584. https://doi.org/10.1021/acs.est.9b06379.
[80] UNEP, 2019. Global Chemicals Outlook II: From Legacies to Innovative Solutions. Launched electronically on 29 April 2019 in Geneva, Switzerland.
[81] Castaño-Ortiz, J.M., Pablo, G.-F., Damià, B., Sara, R.-M., Gil-Solsona, R., 2023. HRMS-based suspect screening of pharmaceuticals and their transformation products in multiple environmental compartments: an alternative to target analysis? J Hazard Mater 132974. https://doi.org/10.1016/j.jhazmat.2023.132974.
[82] Zhang, S., Yin, Q., Wang, S., Yu, X., Feng, M., 2023. Integrated risk assessment framework for transformation products of emerging contaminants: what we know and what we should know. Front Environ Sci Eng 17, 91. https://doi.org/10.1007/s11783-023-1691-3.
[83] Xiong, J., Zhang, Y., Li, Z., 2024. Assessing human exposure to organic contaminants in fish: a modeling approach integrating chemical bioconcentration and food thermal processing. Toxicol Rep 13, 101805. https://doi.org/10.1016/j.toxrep.2024.101805.