Supplementary Materials for

By integrating previously overlooked drivers AI boosts bioaccumulation assessment in fish

HEINZ-R. KÖHLER, REZA AALIZADEH, GABRIELE TREU, THOMAS GRÄFF, KATHARINA PESCHKE, INES PRUTZ, NIKOLAOS S. THOMAIDIS, RITA TRIEBSKORN AND PETER C. VON DER OHE

Corresponding authors: heinz-r.koehler@uni-tuebingen.de (H.-R.K.), r.aalizadeh9@gmail.com (R.A.), peter.vonderohe@uba.de (P.vdO.)

The PDF file includes:

Other Supplementary Materials for this manuscript include the following:

Materials and Methods

Data Collection

The BCF data analyzed were obtained from publicly available scientific literature (e.g. (*1–5*) and the following databases: European Chemical Industry Council: EURAS CEFIC LRI BCF Database (https://cefic-lri.org/toolbox/bcf-database/), National Institute of Technology and Evaluation Japan: NITE 000022750 Dataset 1 (http://www.safe.nite.go.jp/japan/db.html) and NITE 000064114 Dataset 2 (https://www.nite.go.jp>data). All these collected data form the supplementary Datasets S1 and S2. Fig. S2 shows the meta-analysis of the collected logBCF data for (A) 16 different fish species, (B) exposure duration ranging from 2h to 190 days, (C) nominal water concentration ranging from 60 ng/L to 60 mg/L, (D) temperature ranges from 10 to 35 °C and (E) BCF assessment type (kinetic vs steady state). The curated BCF database for several species with experimental conditions is available in the supplementary dataset S3.

Conventional statistics

Data compiled from the public literature and existing databases at the German Environment Agency (UBA) underwent ANOVA analysis and multiple linear regression modelling (MLR) using JMP 16.2.0 (SAS Institute Inc, Cary, NC, USA). MLR was conducted with a standard least square fitting of the model to explain as much of the variance as possible within the BCF data.

Modelling strategy

Neural networks are composed of layers and nodes, as well as inputs and outputs in their architecture. These nodes in each layer are connected by a weighting function applied to the previous input layer. Each layer also contains few nodes to which a function is applied (weighted sum of inputs). When the number of layers between the input and the final output exceeds 1 layer, it is called a Multiple Layer Perceptron (MLP) or DL Neural Network. The image-based DL method requires several supplementary steps/additional procedures to extract and learn from features available in an image. This type of neural network is called C Neural Network (CNN). In other words, the method learns from the different parts of the image and assigns weights and biases on the feature map and, according to a fitting function, it can later recognize or classify the objects in the image into their defined classes. CNN is a specialized form of neural network architecture, which requires connectivity between different nodes and layers to learn from thousands of features (pixels) extracted from an image to improve its classification/regression performance. The process of feature extracting and subsequent learning is controlled by several parameters such as kernel size, filter values, stride, hidden layers, activation functions and number of convolutional layers. For example, one filter parameter controls the size of image segments created from the initial input image. The number of pixels the filter moves each time it operates on the receptive field (kernel size) of an image is called the “stride”. It is important to use a large stride so that the convolution operation scans the entire image, not just part of it. For example, if the size of the receptive field is 7x7, the size of the filter is 3x3 and the stride is 1, then the output of the correlation operation in the 7x7 receptive field will be a 5x5 area. Increasing the stride shrinks the output volume and reduces the computational load, but if the stride is too high, useful information on the image would be lost. In general, the goal is to preserve the maximum amount of information, including the low-level features contained in the input image, especially in the first layers of the network. This is due to the fact that as convolutional operations are being applied, the size of the arrays can shrink even more, and if this happens early in the first layers of the DL architecture, there is a risk of information loss. The first set of connected layers ends with a final layer that has no hidden layers and simply applies the “tanh” or “relu” activation function for further nonlinearity detection. These convolutional layers are connected at each node by a fully connected layer, which is a deep network with n numbers of hidden nodes and again uses an activation function in each node. Usually, the connecting layers between different convolutional layers contain a large number of hidden nodes, while the last connected layer has hidden nodes as many as the number of classes, and one output node for the regression task. It is worth mentioning that each layer also has a pooling layer, which pools the maximum areas of the previous layer, in order to reduce the computational load required to process the data (by finding the dominant features) and also to reduce the noisy activations. Finally, the output of the whole DL is a network that applies

the “softmax” or “linear output activation” functions for the classification and regression cases, respectively. The “softmax” function is suitable for classification problems, because its output can be interpreted as a probability distribution over all possible classes, while the “linear output activation” function provides a quantitative result.

DL individual models

Using the above computational framework, the databases collected for log Pow, log S, biodegradation, logBCF (*Cyprinus carpio*) and acute toxity (*Pimephales promelas*) were modeled for the same input file (“barcode”). The accuracies of the models are shown in Fig. S9.

XGBoost for deriving the BCFpro overall model

To model the BCF across 16 fish species (for their identity see Fig. 4), the log BCF predicted by the above framework in *Cyprinus carpio*, together with the log P_{ow} and log S, and the test conditions (i.e., calculation method (kinetic *vs.* steady-state), uptake duration (in days), and nominal water concentration (in µg/L)) were fused into 16 fingerprints encoding fish species (1 means the measurement was performed on this specific fish species and, hence, 0 means the measurement was not performed on this fish species). Biodegradation and the acute aquatic toxicity for *Pimephales promelas* were not part of the fused model, but these parameters provide additional information on potential degradation products, the adjustment of the duration of exposure time (e.g. for persistent chemicals) or the water concentration used (i.e. to avoid lethal effects). Therefore, these variables were combined to predict the experimental BCF measured in different fish species using the Extreme Gradient Boosting (XGBoost) repressor model (Chen & Guestrin 2016). XGBoost is a parallel tree boosting repressor model that provides more accurate results than other gradient boosting models. The hyperparameters of the XGBoost repressor were optimized by “RandomizedSearchCV” algorithm and a fitness function from “multi:softprob” via “sklearn” and “XGBoost” Python packages. The “leave-one/group-out cross-validation-Q^2” is a widely recognized and commonly used fitness function for regression problems. In this study, CV was set to 3 (the leave one out cross validation was iterated for 3 times for each epoch) and the calculated mean Q^2 was 0.761 (±0.036). Here, the “n_estimators” value was set to 122, which is the number of trees to build, the “learning rate” value was set to 0.05 which provided a step size shrinkage to avoid overfitting, and the “early_stopping_rounds” value was set to 10 (i.e. it stopped the training if the errors of cross validation in the validation set was higher than the training set for 10 consecutive epochs). The “max_depth”, which determines how deep each tree is allowed to expand during the boosting round, was set to 4. The “gamma” value, which determines the threshold for node splitting based on the expected reduction in loss, was set to 0.0091. Finally, “colsample_bytree”, which controls the percentage of features used per tree, was set to 0.730. The regression model was built based on 80% of the initial data (n=3826), and then externally and independently evaluated on the remaining 20% of data (n=767, considering that at least one chemical from each different fish species or assessment test condition was included in the test set). The performance of the XGBoost model is shown in Fig. S10A and compared to simple multiple regression analysis (MLR) in Fig. S9B. Note that the performance of the XGBoost model is superior to the simpler MLR approach. The multi-log BCF model from XGBoost can be interpreted using Shapley values (SHAP). The importance of each variable for the predicted BCF value (Fig. S10C) was calculated using the Shap Python package (*6*). In addition to the log BCF (baseline model via *Cyprinus carpio*), nominal water concentration, log P_{ow}, water solubility, test duration and then fish species fingerprints are important parameters. Note that the test temperature was found to have only a minimal effect on the overall accuracy of the model.

BCFpro Software

A stand-alone software to perform BCF calculations in 16 fish species (BCFpro) is distributed free of charge at: https://www.umweltbundesamt.de/publikationen/BCFpro. It can be used to assist researchers, industry or regulators in assessing BCFs for particular fish species, to screen large numbers of chemicals for their bioaccumulation potential, and to find optimal test conditions in terms of test duration and a nominal/measured water concentration (at start) to be tested. The search space (i.e. BCF at specific conditions) of this evaluation is enormous and very complex, with a massive number of local optima, therefore, meta-heuristics are good

options for optimizing solutions. To optimize test conditions, Grey Wolf Optimizer (GWO) (*7*), a meta-heuristic optimization technique inspired by the leadership hierarchy and hunting mechanism of grey wolves in nature, was applied to the XGBoost model equation. GWO secures to find the optimal experimental conditions for 16 fish species (using chemical concentrations ranging from 10 pg/L to 100 mg/L with a cut size of 1pg/L, and exposure time ranging from 2 hours to 190 days and with a cut step of 1 hour) at which 90% and 50% of the maximum BCF that is possible under realistic lab conditions are expected.

Supplementary Text

Remarks on the correlation between the BCF of fish, the concentration of chemicals in the water, and the exposure time.

Current practice for estimating the bioaccumulation of organic chemicals in fish assumes a constant equilibrium between the aqueous phase (water, aquarium) and the lipid-containing phase (fish body), particularly in the steady-state. Under this equilibrium assumption, organic substances are expected to be distributed according to their lipophilicity, as is the case in the experimental determination of P_{ow} in a flask with two equally sized phases, an aqueous (water) and an organic (octanol) phase. According to this view, the distribution of a given chemical in the aqueous and organic phases (BCF: fish body, P_{ow}: octanol) should result in a chemical-specific constant ratio of the chemical concentrations in these two phases according to the intrinsic lipophilicity of the chemical, which is independent of its initial concentration: although a high initial water concentration will result in higher concentrations in both the organic and aqueous phases than a low initial concentration, once the 'steady state' has been reached, the ratio remains the same. The P_{ow} is therefore calculated as the quotient of the concentration in the octanol and the concentration in the water, and the BCF is calculated as the quotient of the concentration in the fish body and the concentration in the exposure medium. As the solubility of a substance in octanol is generally considered to be a good surrogate for its solubility in the fat (lipid) phase of organisms, current practice in chemical assessment worldwide assumes that the P_{ow} can therefore also be considered a good surrogate for the BCF.

However, this view is based on a number of assumptions that are massively challenged by our study. Firstly, numerous data used for regulatory purposes show that the experimentally determined BCF for a given chemical and fish species can vary - in some cases by more than two orders of magnitude. Secondly, the basic experimental design for determining P_{ow} and BCF is not the same and, as our work shows, thus the P_{ow} of a chemical alone cannot reliably predict its BCF.

We were able to show a negative relationship between the concentration of organic chemicals in the exposure medium and the BCF value in fish, regardless of the identity and lipophilicity of the chemicals (Fig. 1). This relationship was verified by linear regression analysis and confirmed by XGBoost in the AI model, allowing BCFpro to generate three-dimensional concentration-response curves for all chemicals and exposure times (Fig. 6, panel 6).

In the process of P_{ow} determination, where identical volumes of water and octanol are used, the uptake of a chemical into the organic phase reduces the concentration of the chemical in the aqueous phase: For example, if there are 1000 molecules in the system and 990 molecules in the organic phase after 'steady-state' determination, then only 10 molecules remain in the water. The P_{ow} would then be 990 / 10 = 99. If the amount of chemical used was increased by a factor of 10, 9900 of the 10,000 molecules would remain in the octanol and 100 in the water. The P_{ow} would then be 9900 / 100, i.e. also 99. This principle applies to any number of molecules that does not exceed the maximum solubility in water or octanol and is not so small that stochastic processes would be involved.

In contrast, when determining the BCF, the volume ratios between the aqueous and organic (lipid) phases differ by many orders of magnitude, resulting in a clear disproportion in terms of the absolute number of molecules of a test substance that are (or can be) present in the aqueous and organic phases.

An adult zebrafish weighs approximately 200 mg (*8*), i.e. has a volume of 200 µL at an assumed density of ρ=1. The lipid phase in this animal has an even smaller volume: when correcting experimentally determined BCF for the lipid content of fish, a lipid content of 5% (occasionally 6%) is usually assumed, as this corresponds to the average lipid content of fish used in OECD 305, including *Oncorhynchus mykiss*, *Lepomis macrochirus*, *Danio rerio*, *Pimephales promelas* and *Cyprinus carpio* (*9*). The effective lipid phase in zebrafish must therefore be assumed to be approximately 10 µL. In contrast, the volume of the aquarium is disproportionately larger, and a factor of 10^5 to 10^6 seems realistic (for 1-10 litre aquaria).

Thus, if a chemical is taken up by the lipid phase of the fish, the concentration of that chemical in the surrounding water is only minimally reduced - this is shown by countless chemically analysed recovery rates when comparing exposure concentrations before and after an experiment. Here (as in the main text of the present article) 'uptake' refers to net uptake, i.e. the difference between the amount of chemical that has diffused into the fish and the amount of chemical that has diffused out of the fish in the same period of time. Even at low test concentrations in the aqueous medium, in the vast majority of cases, the absolute number of molecules present in the test system exceeds the absorption capacity of the very small lipid volume in the fish. This is even more the case at high test concentrations, but as the external concentration in the water *de facto* hardly decreases during exposure, 'equilibrium' (in the sense of the P_{ow} determination) cannot be assumed. The diffusion pressure exerted on the fish by the virtually unchanged chemical concentration in the water remains similarly high throughout the exposure period, even though high concentrations have already been reached in the lipid phase. This also explains the statistically proven dependence of BCF on exposure time in our study. In our opinion, a 'steady state' in experimental BCF determination is achieved, if at all, only by reaching the maximum lipid solubility in the organic phase and not by 'equilibration' in the sense of P_{ow}.

The maximum lipid solubility of a chemical does not depend on the initial concentration in the test. Ultimately, the maximum possible concentration in the lipid phase (i.e. the saturation concentration) can be reached at both high and low water concentrations if the exposure time is long enough (see Figs. 2 and 3). However, based on theoretical considerations, it is expected that this saturation concentration will be reached earlier at high water concentrations than at low water concentrations.

As explained above, it can be assumed that after prolonged exposure the concentration in the organic phase (fish) will approach the saturation concentration, the level of which is independent of the tested chemical´s concentration in the surrounding water. When calculating the BCF, however, the water concentration (which hardly decreases during the test) is in the denominator of the equation. Therefore, for mathematical reasons alone, high test concentrations are very likely to result in lower BCF values compared to low test concentrations, as shown in our study.

This is illustrated by the following example calculation: Based on the equations given in general form in Fig. 3 in the main text, it is assumed, for example, that the saturation concentration in fish for a defined chemical is 1000 µg/kg. The chemical is tested at three concentrations: one half of the assumed saturation concentration, i.e. at 500 µg/L (x = 0.5; red curve), one tenth of this first test concentration, i.e. at 50 µg/L (x = 0.5 and n_1 = 10; green curve), and one hundredth of this first test concentration, i.e. at 5 µg/L (x = 0.5 and n_2 = 100, blue curve). This results in the bioconcentration factors (BCFs) shown in Fig. S6. For both time points considered, t_1 and t_2, an inverse relationship between test concentrations and resulting BCFs is evident. Conversely, for all of the three exposure concentrations considered, longer exposure times of the given chemical result in gradually higher BCFs.

These trends are supposedly consistent for the majority of chemicals evaluated. However, it might be conceivable that in the case of very lipophilic substances (with log P_{ow} values above 6), which are tested at very low concentrations, accumulation even in an extremely small-volume lipid phase will result in a measurable reduction in concentration in the aqueous phase. However, even for these specific chemicals, elevated concentrations in the aqueous phase (exceeding saturation thresholds) consistently will lead to reduced bioconcentration factors (BCFs), mirroring the inverse relationship between concentration and BCF described above. Applied to the natural situation in flowing waters, where, for example, contaminated water

may be continuously supplied, e.g. from a point source, the notion of a fish completely absorbing nearly all molecules of a highly lipophilic substance from the water is also extremely unrealistic.

Although the results shown can be explained mathematically by the relationships described above, toxicodynamic processes within the body's own metabolism certainly also contribute to the level of the BCF. For example, very low internal concentrations of a chemical, i.e. in the case of low external concentrations, may fall below the threshold for inducing a metabolic reaction. In this case, below this threshold, there would be no induction of metabolic enzymes (e.g. the cytochrome P450 family) and thus no increased elimination of the chemical. At higher concentrations, however, the induction of biotransformation enzyme expression may result in disproportionately greater elimination of the chemical. This effect, in addition to the mathematical implications explained, may also reduce the BCF, if active excretion is rapid and the parent compound does not diffuse back to a great extent to balance the internal concentration again. On the other hand, depending on the toxicity of the chemical in question, however, very high internal concentrations may lead to an overload of metabolic enzymes. Several Phases I and II enzymes are described within the metabolization and processes in fish (notably in liver, gills and tissue) which convert hydrophobic compounds into more hydrophilic metabolites with reduced membrane permeability. In phase III, the hydrophilic metabolites are then actively exported from the cell by transmembrane transport proteins. In fish, active transporters expressed in membranes of tissues with excretory (e.g., liver, kidney), absorptive (e.g., intestine), and barrier (e.g., skin, blood–brain, blood–eye and blood–gonad barriers) were described (reviewed by (*10*)). For instance, P-glycoprotein (P-gp) transporters eliminate chemicals from the body through excretory fluids. P-gp transporters can be induced or inhibited by chemicals depending on the timing and concentration of exposure (*10*, *11*). However, there is insufficient data on active transport rates in fish and it remains unknown whether they are comparatively faster than passive transport rates and under which circumstances these differences become crucial for bioaccumulation assessment (*5*). The BCFpro model currently does not account for these biological processes, which introduces inherent uncertainties and limits the absolute accuracy of *in silico* predictions. However, despite these limitations, our system achieves a prediction reliability of up to 90%, which can be considered a significant advancement in the field.

Detailed remarks on the use of P_{ow} as screening criterion

Octanol is used as a surrogate for fish lipids as screening criterion for bioaccumulation in chemical legislation. At log P_{ow} values between 4 and 5, Log BCF increases linearly with log P_{ow} if the substance is absorbed at the same rate and is not biotransformed (*12*). This linear relationship is the basis for the B screening threshold value of log P_{ow} >3, indicating that the substance may bioaccumulate to a significant degree (*13*). However, this approach has some limitations, especially for hydrogen-bond donors, polar or ionic compounds, which have stronger interactions with membrane lipids compared to storage lipids or octanol (*12*, *14–16*). Accordingly, we demonstrated in the current study that P_{ow} explained only 45,5% of the variation in reported BCFs and that exposure duration and water concentration have a significant impact also. Thus, these two descriptors and other parameters should be required to be factored in as well when screening for bioaccumulation in the future. This means regulatory bodies continue to rely on P_{ow} as a primary factor in determining the necessity for subsequent in vivo bioaccumulation testing, despite it being increasingly recognized as an insufficient standalone predictor of bioaccumulation potential. At screening level, we recommend using the BCFpro model which allows us to estimate a more realistic worst-case BCF. Particularly DL-based models like the BCFpro have the potential to significantly enhance bioaccumulation assessment at the screening level in the future, offering rapid, cost-effective, and animal-free alternatives to traditional methods. As these technologies continue to evolve, they are likely to play an increasingly important role in chemical safety assessment and environmental protection.

Remarks on limitations and advances of the study

Data quality of experimental BCF studies

Notably, measured bioaccumulation data are quite limited compared with the breadth of the chemicals currently used in commerce (*17*) or experimental studies on ecotoxicological endpoints, which are not publicly

accessible due to confidential business information (*18*). Due to this general lack of measured data for bioaccumulation, and time and budget restrictions, bioaccumulation assessments must rely on several estimation methods or fish studies at significant costs. Uncertainty is inherent in PBT assessments, whether the supporting data are measured or estimated, but uncertainty increases when estimation methods are applied to chemical classes for which empirical data is sparse (*17*). Furthermore, it needs to be considered that studies are not necessarily more reliable because they were determined experimentally according to, for example, OECD and GLP standards, which has been shown for the determination of P_{ow} (*19*).

Especially for superhydrophobic substances ($P_{ow} > 6$), the determination of the bioaccumulation potential can be difficult or lead to error-prone BCF values (*20*) due to several issues: difficulties in measuring the true aqueous concentration due to sorption of the substances to particulate and dissolved (organic) matter; adsorption of the test substance to glass walls or other materials; unstable concentration of the test substance in water and thus highly fluctuating exposure conditions; volatilization; testing at concentrations clearly above the water solubility of the test substance, normally via the inclusion of dispersants or vehicles which would lead to an underestimation of the BCF; and determination of a BCF as the ratio between the concentration in fish and in water, but under non-steady state conditions (*13*). To address these limitations, various changes have been incorporated in the new Technical Guideline 305 in 2012, including the recommendation for a dietary exposure test for hydrophobic substances ($\log P_{ow} > 5$) with a solubility less than approximately 0.01–0.1 mg/L (*21*). Despite this development and the acknowledged shortcomings, many BCF values for hydrophobic substances from old OECD 305 studies are still used and accepted as "correct". As part of the experimental data (35,4% of the 965 chemicals assessed) in the current work were recorded before the guidance update in 2012, it cannot be excluded that some of the BCF values, notably for hydrophobic chemicals, were underestimated in the original studies, which might have affected the modelling.

Total organic carbon and freely dissolved concentration

The freely dissolved concentration refers to the amount of a chemical substance that is actually dissolved in the water and available for uptake by organisms (*22*) and is therefore considered the bioavailable fraction of the chemical in the aquatic environment. Organic matter, quantified as total organic concentration (TOC) and as dissolved organic carbon (DOC), can have a significant effect on the amount of freely dissolved test substance in flow-through BCF fish tests, especially for highly hydrophobic substances (*23*, *24*). The total water concentration includes both the freely dissolved chemical and the fraction bound to particulate matter or dissolved organic carbon in the water. Sorption of the test substance to organic matter can reduce its bioavailability and therefore lead to a significant underestimation of the BCF (*2*). Thus, high concentrations of particulate and dissolved organic carbon in the water of the test vessel reduce the true BCF when calculated from total water concentrations (*9*, *23*). Therefore, the concentration of TOC in the test vessels in an OECD 305 study should not exceed the concentration of organic carbon originating from the test substance (and solubilising agents, if used) by more than 10 mg L^{-1} (*9*). However, significantly reduced bioavailability of hydrophobic test substances may occur even in the presence of rather low concentrations of TOC (i.e. < 10 mg L^{-1}) (*24*). As a result of reduced bioavailability, high TOC concentrations typically result in lower measured BCF values compared to what would be observed in water with low TOC, as the BCF is often calculated based on total water concentrations, not just the freely dissolved fraction (*23*). However, the determination of freely dissolved substance concentrations and the calculation of the BCF value based on them is not mandatory in Europe (*23*).

As most experimental studies did not report the TOC content of the test system or the freely dissolved water concentration, their effects on the BCF could not be systematically investigated. However, based on the above relationship, we assume that the true bioaccumulation potentials of the chemicals may be even higher than estimated by the BCFpro model, since we used the measured total water concentration and could not consider the freely dissolved concentration. We recommend calculating the BCF based on the freely dissolved water concentration using appropriate extraction methods like solid-phase microextraction (SPME) (*23*, *24*). Furthermore, we emphasise that future studies should minimize the TOC content, explicitly report the TOC concentration and justify all experimental decisions and conditions that will allow the effect of TOC to be considered in future modelling approaches as more data become available.

Correction for growth

The OECD 305 test guideline outlines a method for growth correction in BCF studies to account for the effect of fish growth on chemical concentrations in tissues. This correction is necessary because as fish grow during the experiment, the concentration of the test substance in the fish may decrease due to growth dilution, even if no actual elimination of the substance occurs (*25*). On an internal concentration basis, growth is considered a 'biomass-based dilution' or pseudo-elimination effect in the fish, while the magnitude of growth-related underestimation can be significant (*26*). Studies have shown that failure to account for growth can result in BCF values that are significantly lower than the true bioconcentration potential of a chemical (*25*, *26*). This is particularly important for slowly excreted substances and for studies using rapidly growing fish. Simple fish growth models can be used to estimate the growth rate constant and its uncertainty, which can be used to correct kinetic BCF values but not steady-state BCF values for growth. The OECD 305 guideline recommends a growth correction method in which the growth rate of the fish is subtracted from the measured depuration rate constant to obtain a growth-corrected depuration rate constant. So, proper accounting for fish growth is critical to the accurate determination of BCFs. Failure to do so may result in a significant underestimation of the bioaccumulation potential of a chemical, which could have important implications for environmental risk assessment and regulatory decisions. Nevertheless, for reasons of availability, cost and ease of experimentation, juvenile fish are often used, but the BCFs were not corrected for growth, nor were the values reported for most of the substances in the current work which may be one reason for the underestimation of some of the estimated kinetic BCF values. Unfortunately, because raw data on fish growth during the test phase are required but not available, we could not correct the kinetic BCFs for growth.

ML models for bioaccumulation assessment and regulatory perspectives

Reliable computational tools can predict ecotoxicological endpoints and the bioaccumulation factor (i.e. BCF) of organic chemicals to reduce the need for animal testing. To be considered viable for this application, model predictions should be consistent and explainable over a wide chemical and taxonomic space. These models struggle with limited experimental data, varying reliability and heterogeneity, and imbalances in chemical space coverage (*27–29*), but ignore the test conditions and their impact on the reported measurements. Typically based on single algorithms, traditional QSAR models often lack interpretability and rely heavily on physicochemical descriptors such as log P_{ow}, neglecting critical factors such as chemical structural features and test conditions—including water chemistry, test concentration, fish species and biodiversity, duration, and temperature (*2*, *17*, *30–41*). This oversight can have significant implications because test conditions can affect BCF measurements more than traditional parameters, resulting in low predictive accuracy and challenges in integrating data across fish species (*1*, *17*). For a single chemical with multiple measurements, averaging the logBCF values across various fish species has resulted in an increased prediction error (*42*). Properly documented measurement metadata is key to identifying high-quality data for training ML model training, for example, to include only datasets with consistent experimental conditions following standardized protocols (*28*). A handful of previous models, their full accuracy (both training and validation set) as well as study design have been reviewed and can be found in Table 1 and Supplementary Database S4). To overcome these limitations, we developed BCFpro, a machine learning model designed to improve BCF prediction by directly incorporating chemical structural features and accounting for test conditions and fish species sensitivity. BCFpro leverages metadata, such as species characteristics and experimental variables, to significantly improve the accuracy and reliability of BCF predictions. Unlike traditional QSAR models, BCFpro does not rely solely on physicochemical descriptors but integrates a broader range of factors, enabling BCF estimation under different test conditions and across multiple fish species. This versatility is critical because test conditions can have a disproportionate effect on BCF values (*43*). By accounting for these variables, BCFpro provides a more robust and comprehensive approach to predicting bioaccumulation potential. One of the notable strengths of BCFpro is its ability to provide a holistic understanding of bioaccumulation potential. It considers persistence and acute toxicity, as well as 16 fish species, providing a more comprehensive assessment than traditional QSAR models, which often just predict a single BCF. This multi-species approach is supported by metadata integration, which improves prediction consistency and allows for species-specific

insights (*14*). Traditional models, limited by their reliance on uniform descriptors, often fail to capture such variability, making BCFpro a superior tool for comprehensive bioaccumulation assessments. Recent work by Wassenaar and co-workers (*1*) on a BCF dataset similar to ours (NITE dataset, Arnot and Gobas, 2006) showed high variation within BCF values for individual substances (mean SD of log BCF = 0.21), which calls into question the robustness of the current bioaccumulation assessments. Considering that SD, for example, the 95% confidence range of a hypothetical BCF value of 2500 ('bioaccumulative') would range from 953 ('not bioaccumulative') to 6561 ('very bioaccumulative').

BCFpro also excels in regulatory applications by providing estimates under specific conditions and insight into worst-case scenarios, such as BCF50 and BCF90 values using the Gray Wolf optimizer coupled with the XGBoost model. These metrics, which indicate test conditions where higher bioaccumulation may occur, are critical for strengthening regulatory assessments and chemical management (*43*, *44*). Traditional QSAR models rarely account for such scenarios, limiting their usefulness in decision making. BCFpro's ability to highlight these extremes, combined with its consideration of test conditions, positions it as an invaluable tool for sustainable chemical design and regulatory compliance.

Fig. S1.

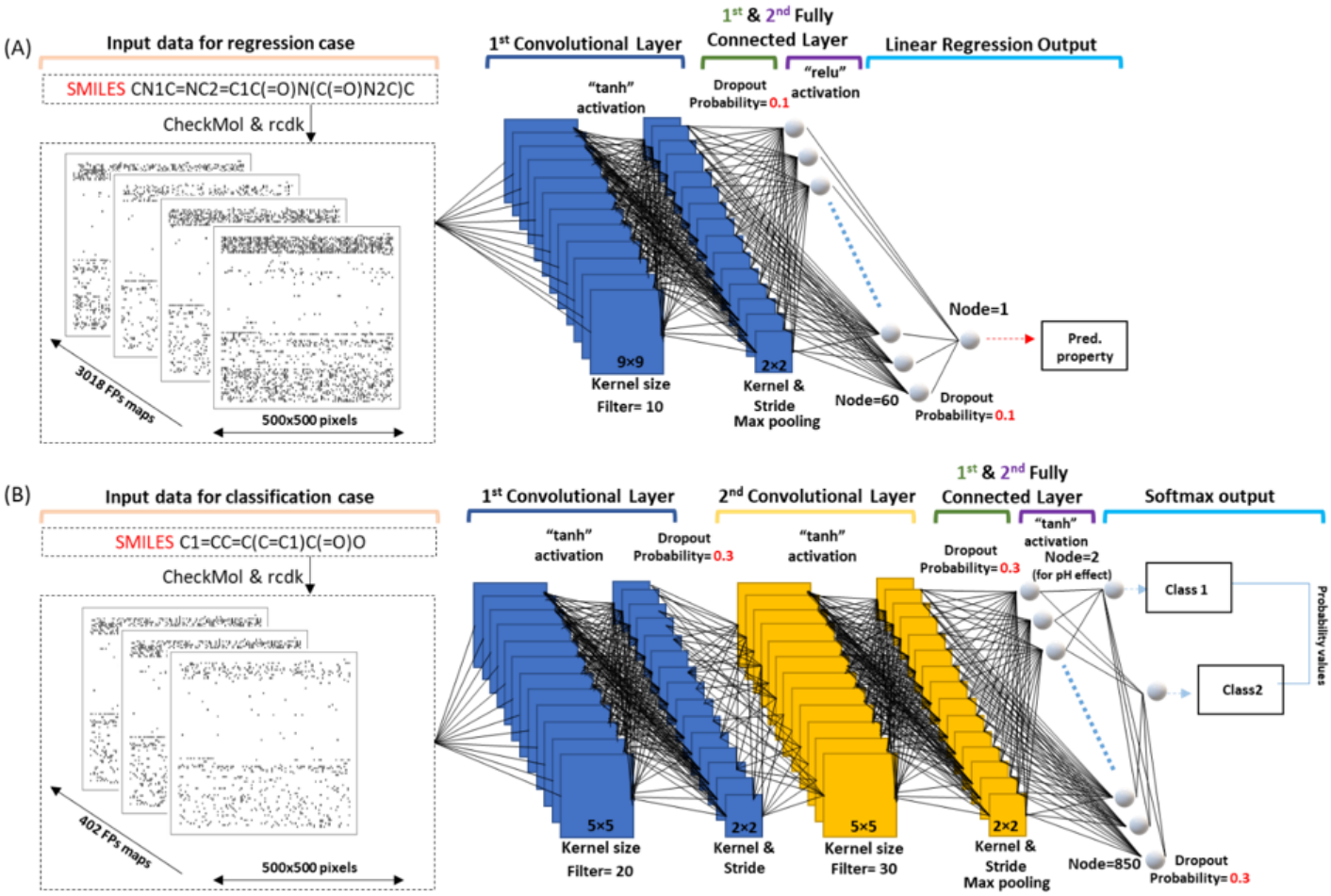


Fig. S1: Convolutional neural network deep learner architecture for modelling of **A**: regression cases, such as $logP_{ow}$, logS, LC_{50} and BCF, as well as **B**: classifications (such as "biodegradable" or "not biodegradable"), figure adapted from (*6*).

Fig. S2.

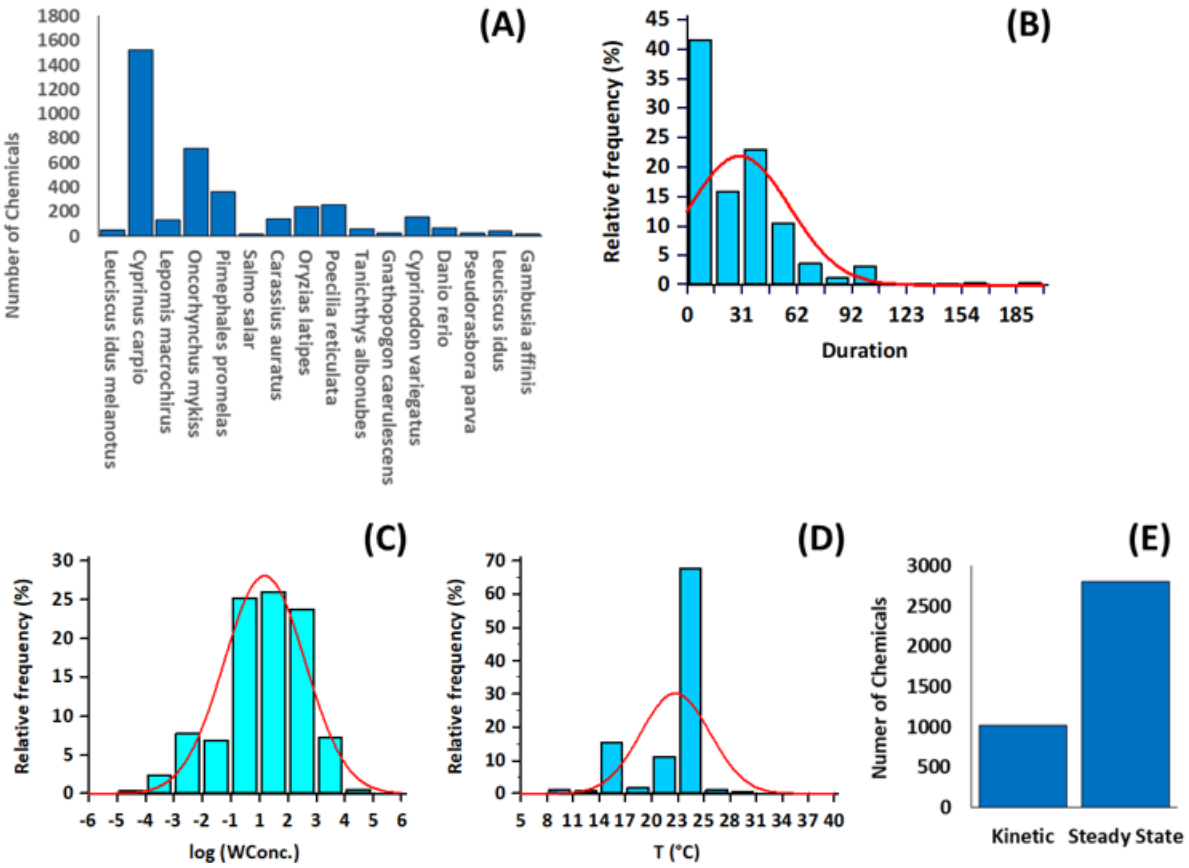


Fig. S2: Meta-data analysis for collected logBCF values for **A**: 16 species; **B**: duration of assessment in days; **C**: nominal water concentration; **D**: temperature of BCF assessment and E: type of BCF assessment. Red lines symbolize the type of data distribution.

Fig. S3.

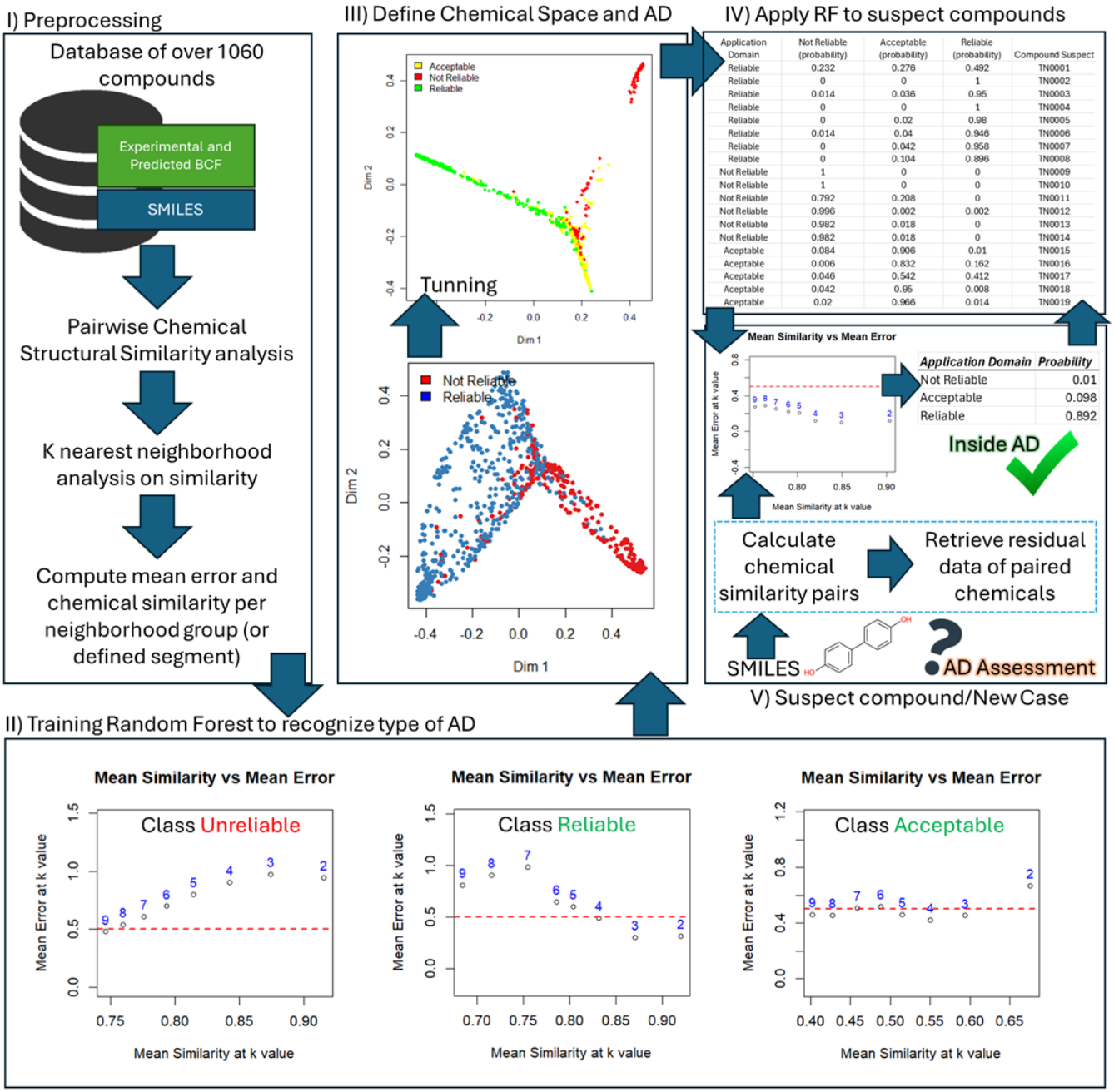


Fig. S3: Proposed procedure for evaluating the application domain (AD) within the logBCF model. The AD is defined as a function of the trend pattern between chemical similarity and residuals. Residual analysis is performed on paired chemicals that are found, while chemical similarity is calculated using the provided SMILES. Consequently, a supervised classification model tuned for AD assessment does not necessitate the experimental observation of a suspect compound.

Fig. S4.

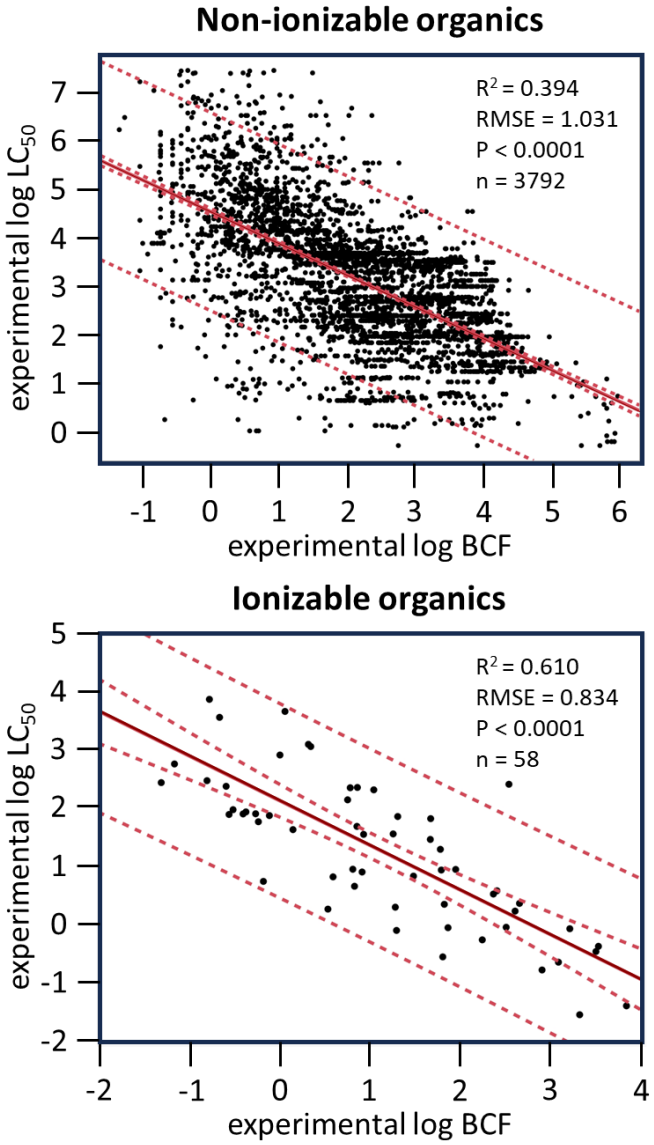


Fig. S4: Experimentally generated log BCF data *vs.* log LC_{50} of corresponding chemicals, own compilation of data from public sources (non-ionizable organics, top) and data from (*3*) (ionizable organics, bottom). Linear regression line (red, solid) and its 95% confidence intervals (red dashed, inner interval) plus 95% confidence intervals for the data point distribution (red dashed, outer interval). Highly significant (P<0.0001) negative correlation (ANOVA) of the two parameters. R^2: Bravais-Pearson variance. RMSE: root mean square error. n: number of data points.

Fig. S5.

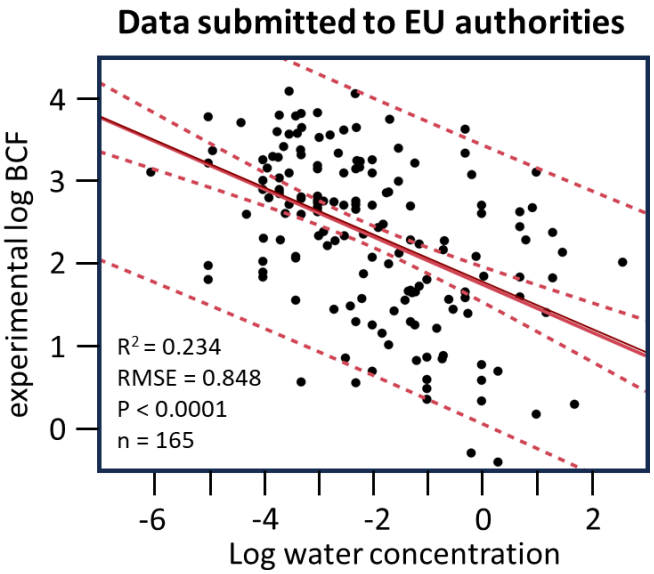


Fig. S5: Experimentally generated regulatory log BCF data *vs.* log concentration in the water. Confidential data submitted to the European Chemicals Agency (ECHA), the European Food Safety Authority (EFSA) and the European Medicines Agency (EMA) that were handled by the German Environment Agency (UBA) for regulatory purposes. 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. RMSE: root mean square error. n: number of data points.

Fig. S6.

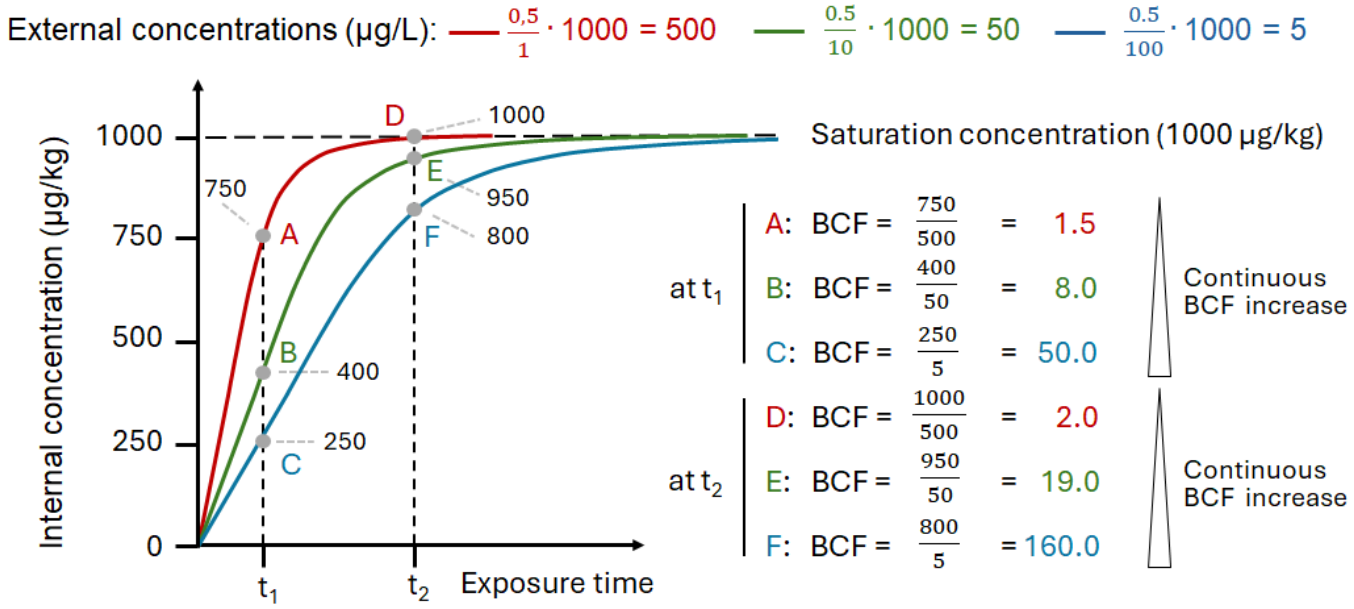


Fig. S6: Accumulation kinetics of a given organic chemical as a function of exposure time. Example calculation to illustrate the mathematical relationships shown in general form in Fig. 2. Exemplary assumption of the following numerical values for the variables s = 1000, x = 0.5, n_1 = 10 and n_2 = 100. Explanation in the main text.

Fig. S7.

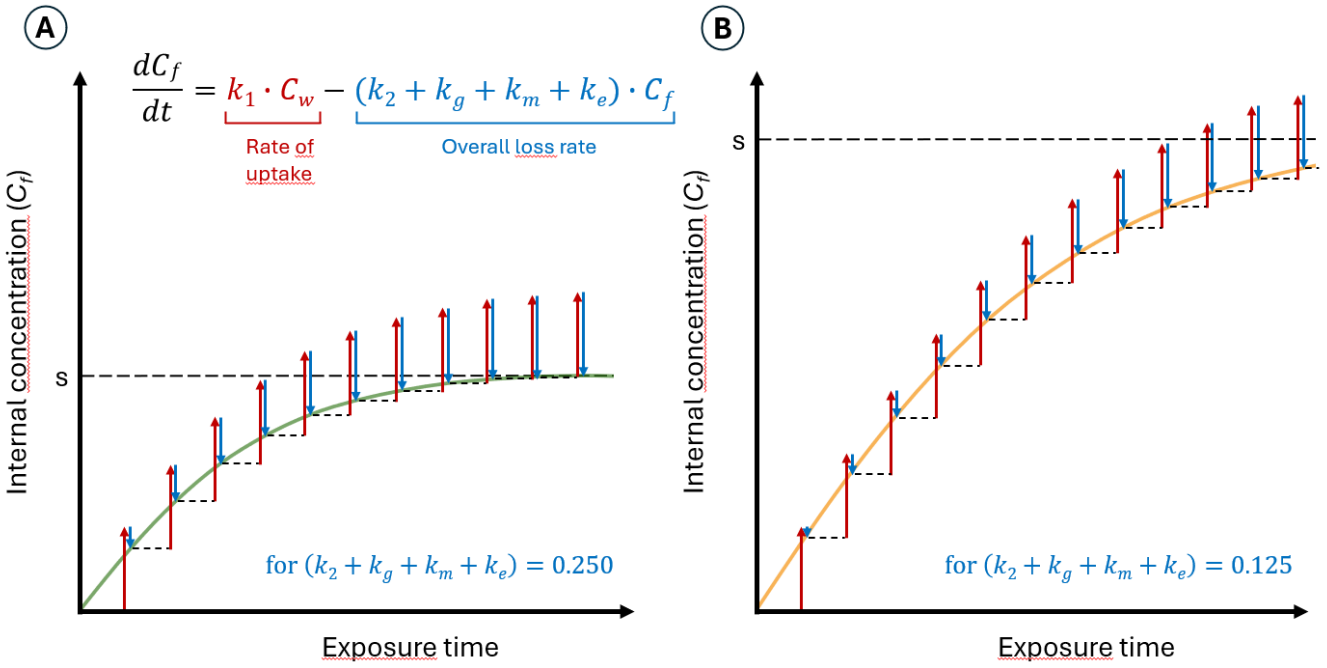
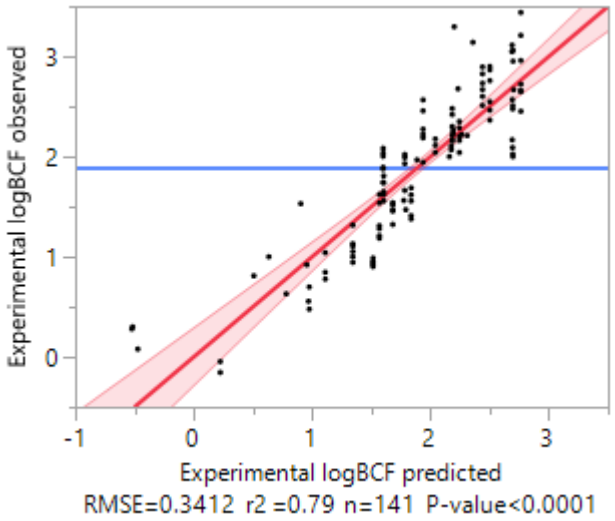


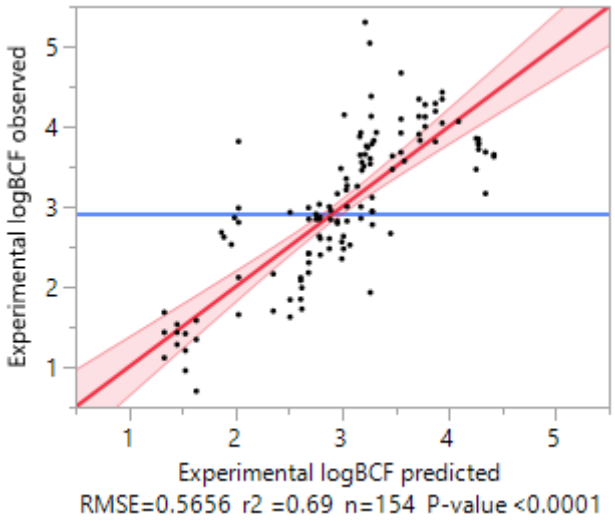
Fig. S7: Accumulation kinetics of a given organic chemical as a function of exposure time for fish with low (**A**, green line) and high (**B**, orange line) lipid content. According to the OECD 305 test guideline (OECD 2017) the general fish bioaccumulation model is described by the uptake rate (in red) and the overall loss rate (in blue) following the differential equation given in the figure (Sijm et al. 1993) where 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). Red and blue arrows visualize uptake and loss of the chemical from the body of the fish at given times. For a given chemical with a constant concentration in the water, the uptake rate k_1 per time unit remains the same over the entire exposure time (red arrows of equal length) because the uptake of the chemical into the fish does not result in a significant reduction of its concentration in the water. In fish with a low fat content (A), the volume of the aqueous phase within the fish body (blood, ascites, cytoplasm) is disproportionately higher than the volume of the lipid phase (adipose tissue, membranes). Consequently, there is also a comparatively rather high absolute number of molecules of the chemical in the aqueous phase of the fish body and potentially available for depuration. The depuration constant is comparatively high (e.g. $k_2 = 0.25$, arbitrarily assumed), which leads to comparatively high total loss rates (blue arrows). After a certain time, the accumulation of the chemical reaches the saturation level (s) in the steady-state phase. In fish with a higher fat content (B), k_1 remains the same, however, the depuration constant k_2 is lower (e.g. $k_2 = 0.125$, arbitrarily assumed, half as high as in (A)), because the elimination of the organics can only take place via the water-soluble part of the chemicals and the largest part of the chemical, corresponding to the P_{ow}, is located in the large-volume lipid/fat phase of the fish. It is clear that the saturation level (s) in this example (with only half the k_2 compared to (A)) is twice as high as in (A) which would therefore also apply to the resulting BCF. Hence, the saturation level in the steady state is only reached after a longer exposure time, compared to (A).

Fig. S8.

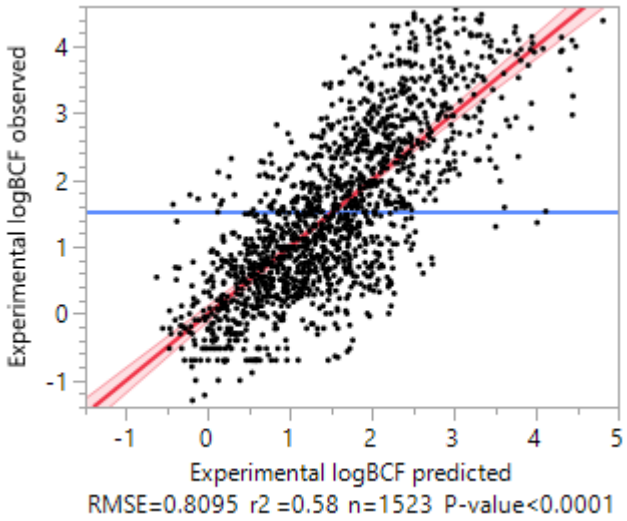
Carassius auratus



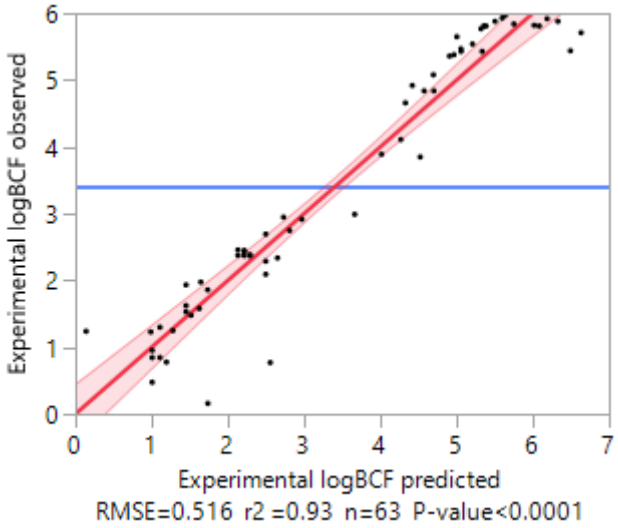
Cyprinodon variegatus



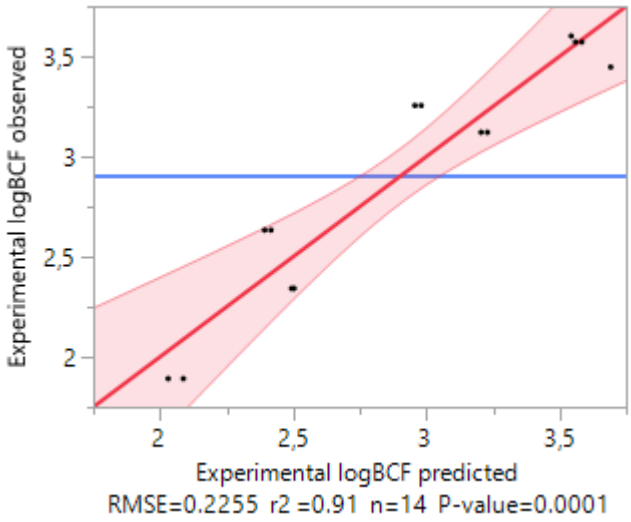
Cyprinus carpio



Danio rerio



Gambusia affinis



Lepomis macrochirus

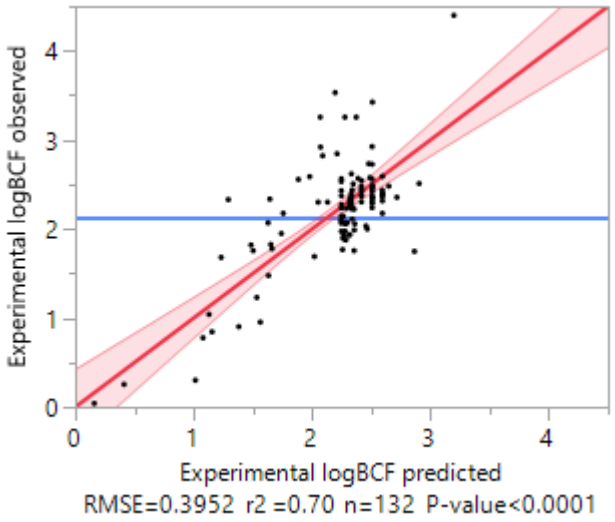
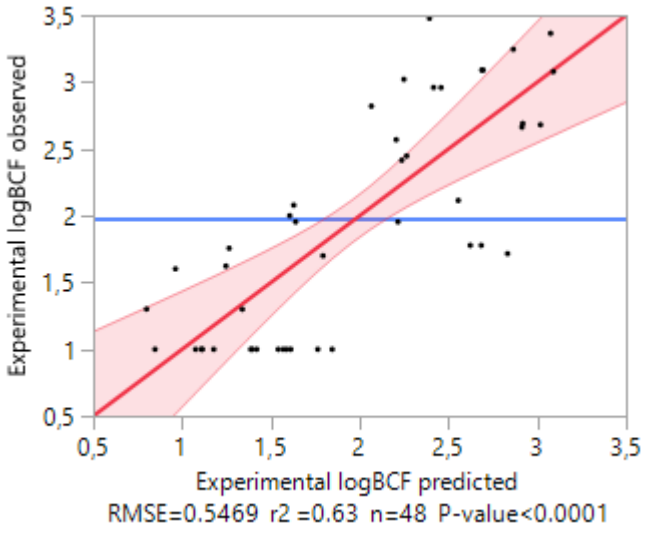
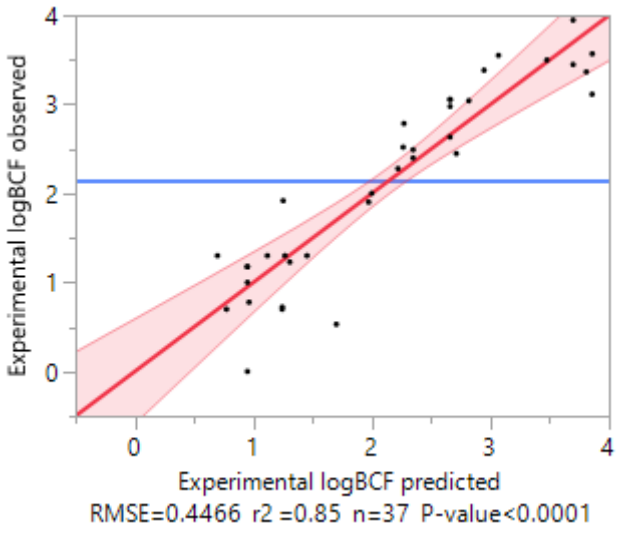


Fig. S8 (continued).

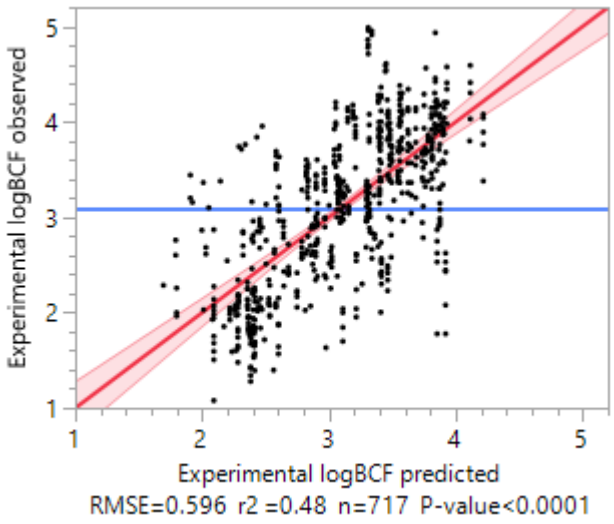
Leuciscus idus melanotus



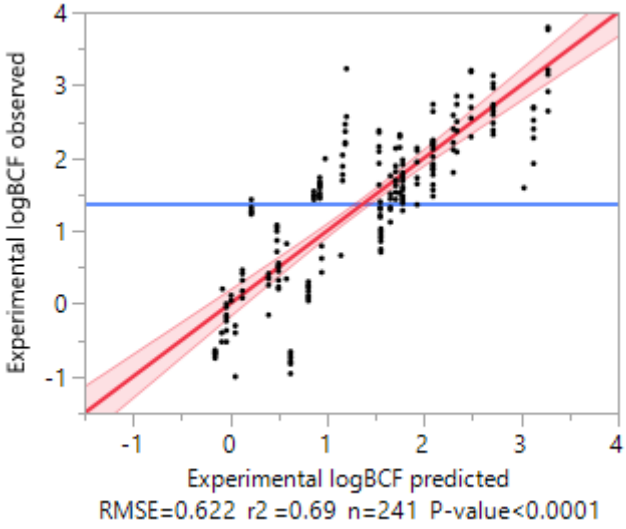
Leuciscus idus



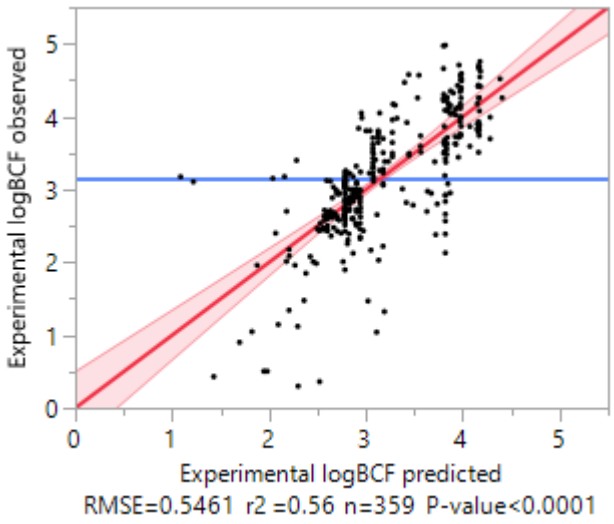
Oncorhynchus mykiss



Oryzias latipes



Pimephales promelas



Poecilia reticulata

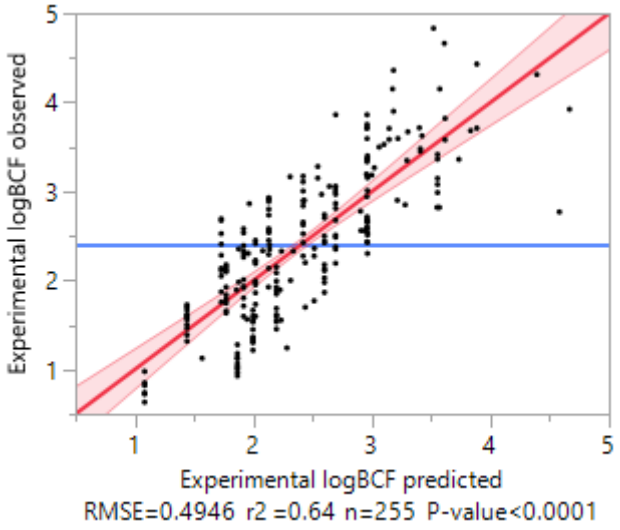


Fig. S8 (continued).

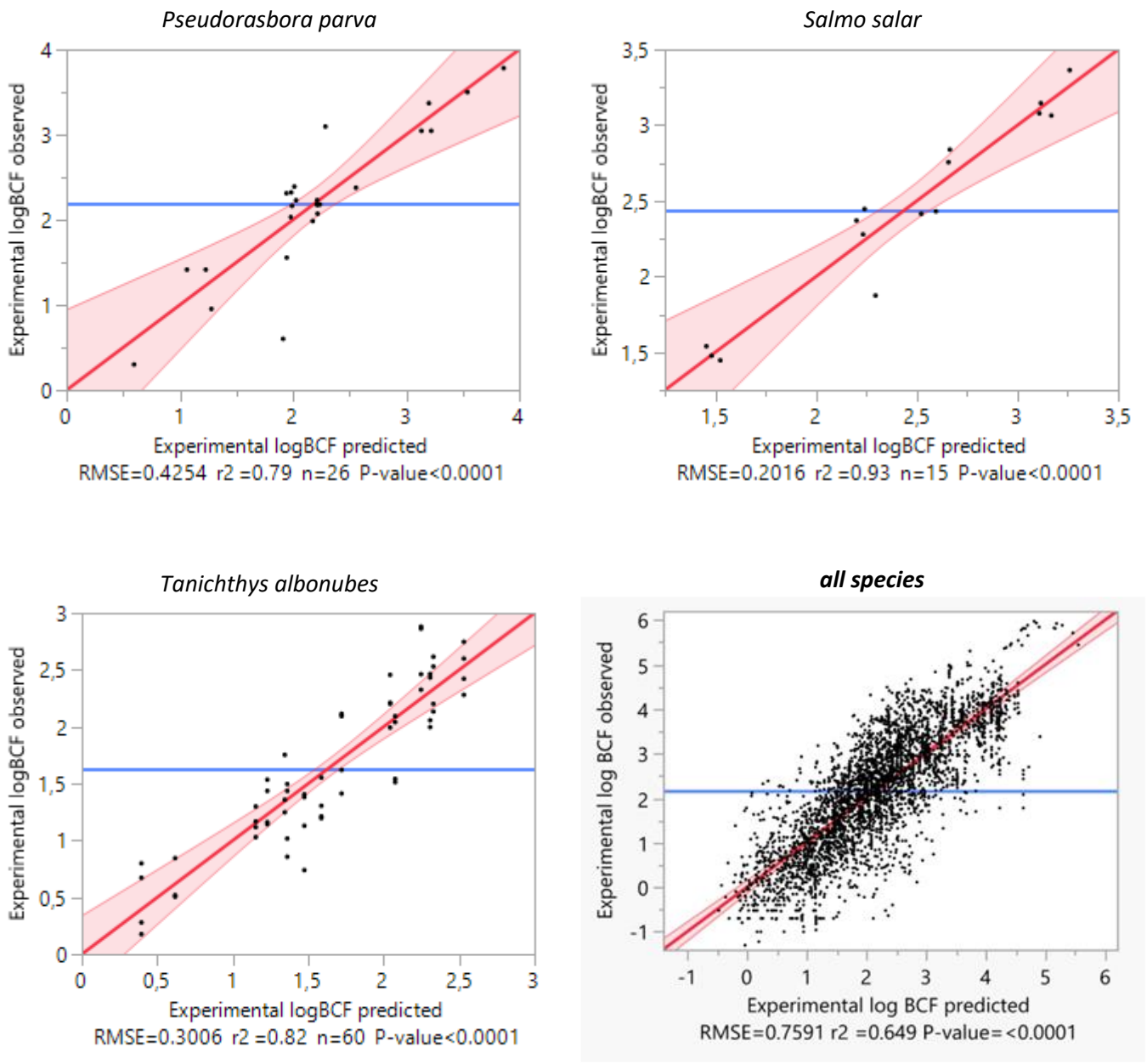


Fig. S8: Species-specific multiple regression modeling (MRM) of experimental logBCF as a function of the predictors $logP_{ow}$, log water concentration of tested chemicals, log water solubility of tested chemicals, water temperature and exposure duration. MRM-predicted logBCF (x-axis) *vs.* experimental BCF (y-axis). Calculations were performed using data from 15 different fish species: *Carassius auratus*, *Cyprinodon variegatus*, *Cyprinus carpio*, *Danio rerio*, *Gambusia affinis*, *Lepomis macrochirus*, *Leuciscus idus melanotus*, *Leuciscus idus*, *Oncorhynchus mykiss*, *Oryzias latipes*, *Pimephales promelas*, *Poecilia reticulata*. There were too few data points for a meaningful regression analysis for *Jordanella floridae*. Bottom right, shaded in gray: MRM for the complete dataset comprising all species using the above-mentioned predictors plus the species as variables. Regression slopes are shown in bold red. Light red areas indicate the 95% confidence interval. The overall mean is represented by the blue horizontal line. RMSE: root mean square error, r^2: coefficient of determination, n: number of observations, P-value: probability value.

Fig. S9.

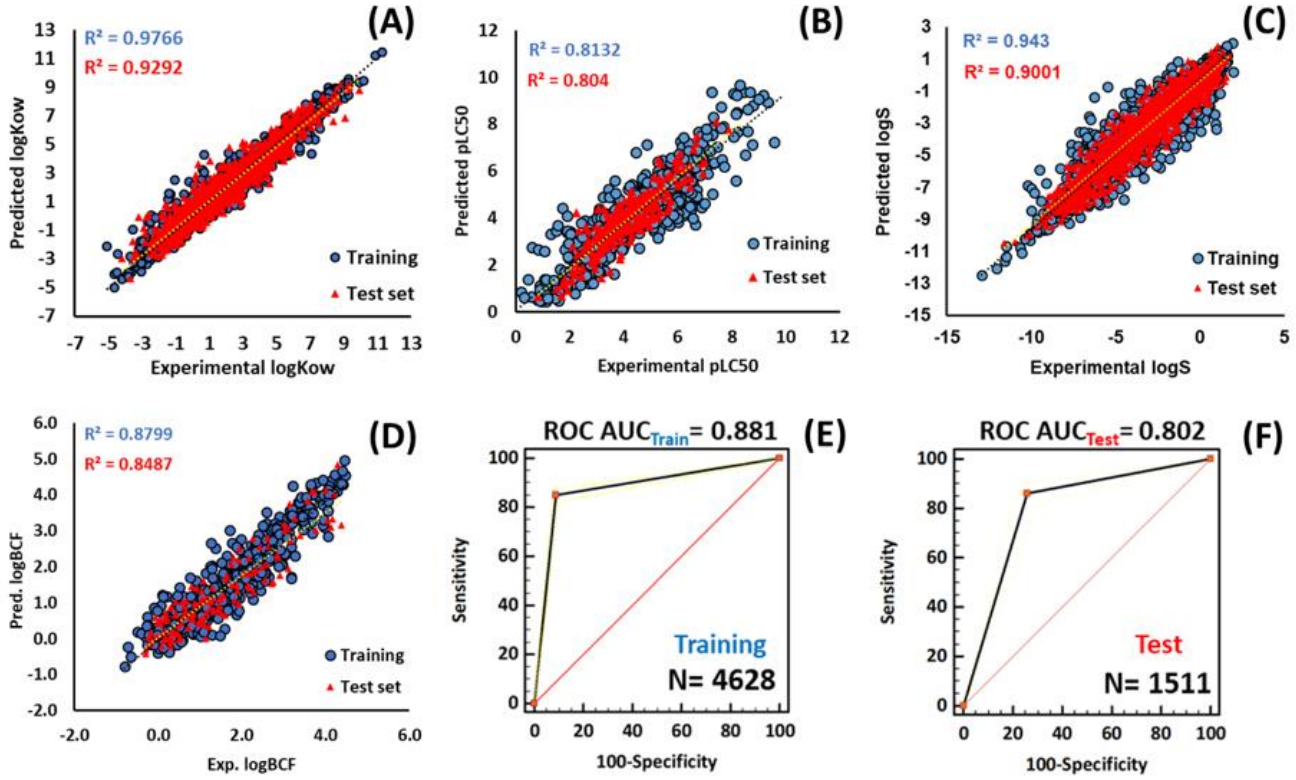


Fig. S9: The correlation plots between experimental and predicted physicochemical properties and biological endpoints used in this study. **A**: Model built for log Pow, n = 13886 (n = 2813 (test 20%, red triangles), n = 11073 (training 80%, blue dots)). **B**: Aquatic acute toxicity (LC_{50}) model built for *Pimephales promelas*, n = 1225 (n = 217 (test 18%, red triangles), n = 1008 (training 82%, blue dots)). **C**: Water solubility model denoted as log S, n = 9768 (n = 1932 (test 20%, red triangles), n = 7836 (training 80%, blue dots)). **D**: DL model for the bioconcentration factor (log BCF) in *Cyprinus carpio*, n = 790 (n = 131 (test 17%, red triangles), n = 659 (training 83%, blue dots)). **E** and **F** show the receiver operating characteristic (ROC) curves of the model predicting aerobic ready and inherent biodegradation of organic chemicals in water, n = 6139 (n = 1511 (test 25%), n = 4628 (training 75%)).

Fig. S10.

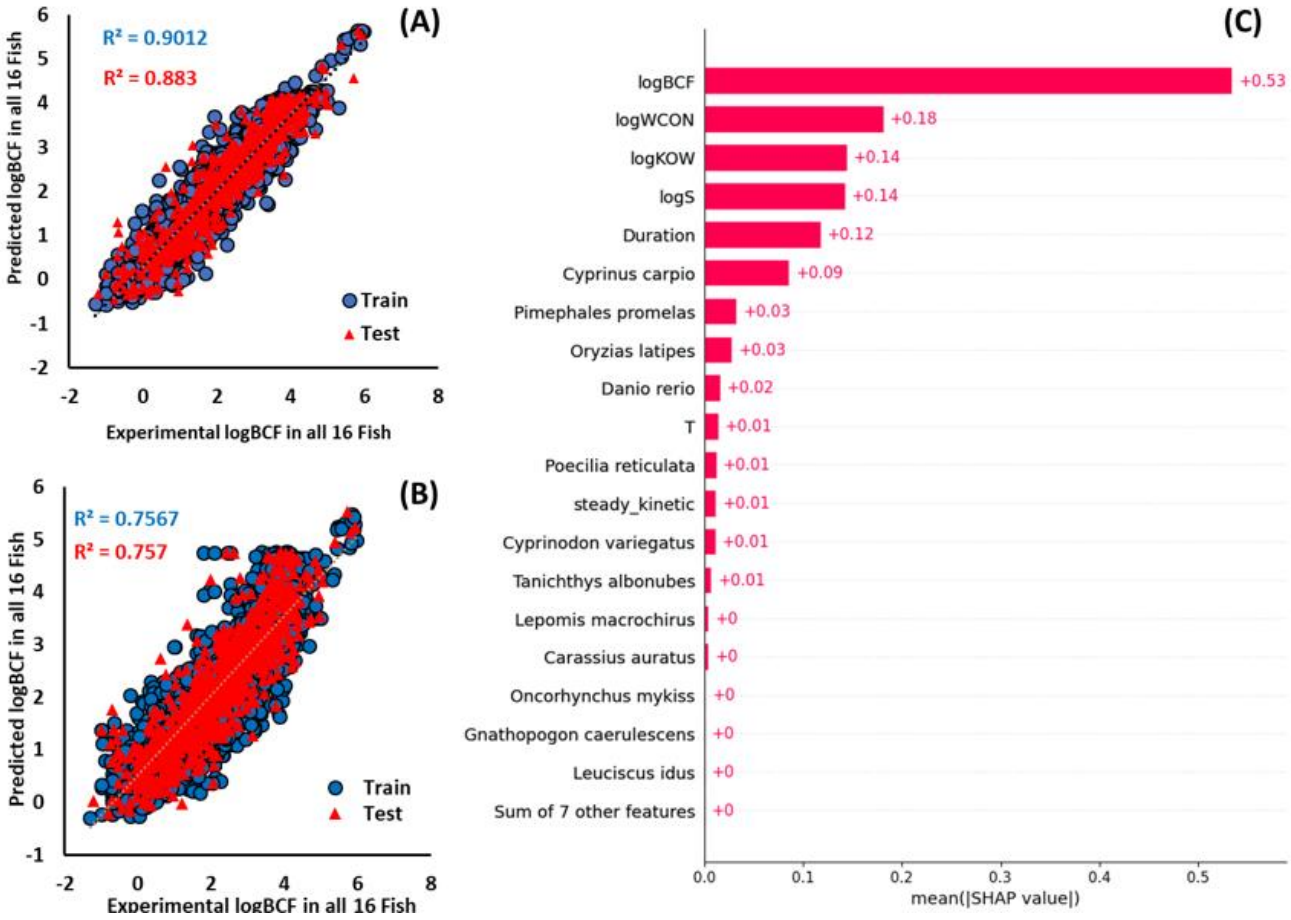


Fig. S10: Correlation plot between experimental and predicted log BCF for 16 fish species using **A**: XGBoost and **B**: the MLR approach. **C** shows the SHAP values (variable importance) for variables used in the structure of the XGBoost regression model. Apart from log BCF (baseline model for *Cyprinus carpio*), nominal water concentration (logWCON), $logP_{ow}$ (logKOW), water solubility (logS), test duration and fish species fingerprints are important parameters. The overall SHAP analysis clearly shows the significant contributions of the water concentration parameter, designated "logWCON", with a dimensionless contribution of +0.18, and the exposure time parameter, designated "Duration", with a dimensionless contribution of +0.12, to the overall performance of the model. While the most significant contribution comes from the DL model for *Cyprinus carpio* (+0.53), the contribution of concentration and exposure time is equal to that of other commonly used parameters such as log Pow (designated "logKOW") and water solubility (designated "logS"), both of which are at +0.14.

Fig. S11.

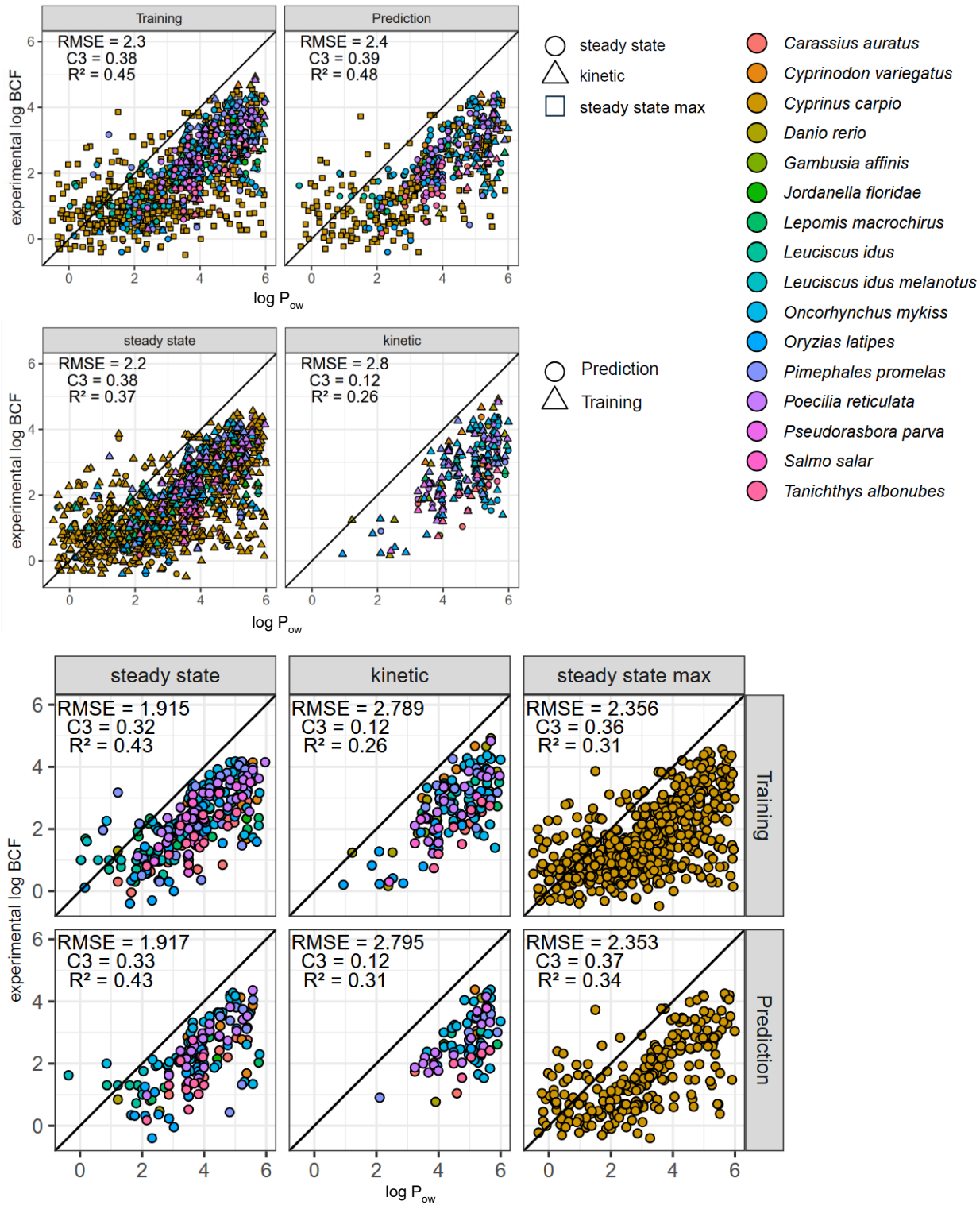


Fig. S11: Experimentally generated log BCF data *vs.* log P_{ow} of corresponding chemicals, own compilation of data from public sources. Separate presentation of data presumably collected at steady state, collected during the kinetics of progressive accumulation or maximum values presumably collected at steady state conditions (circles, triangles and squares, top plots) as well as data used for AI training and those used exclusively for comparison with AI-generated predictions (triangles and circles, middle plots). The bottom six plots show the separation for both criteria (i.e. steady-state and kinetic). The different colors of the symbols refer to the fish species tested. RMSE: root mean square error. C3: concordance correlation coefficient as a measure of how well the data set fits the 1:1 line (black bisectors). R^2: Bravais-Pearson variance. It becomes evident that log P_{ow} alone does not adequately represent log BCF.

Fig. S12.

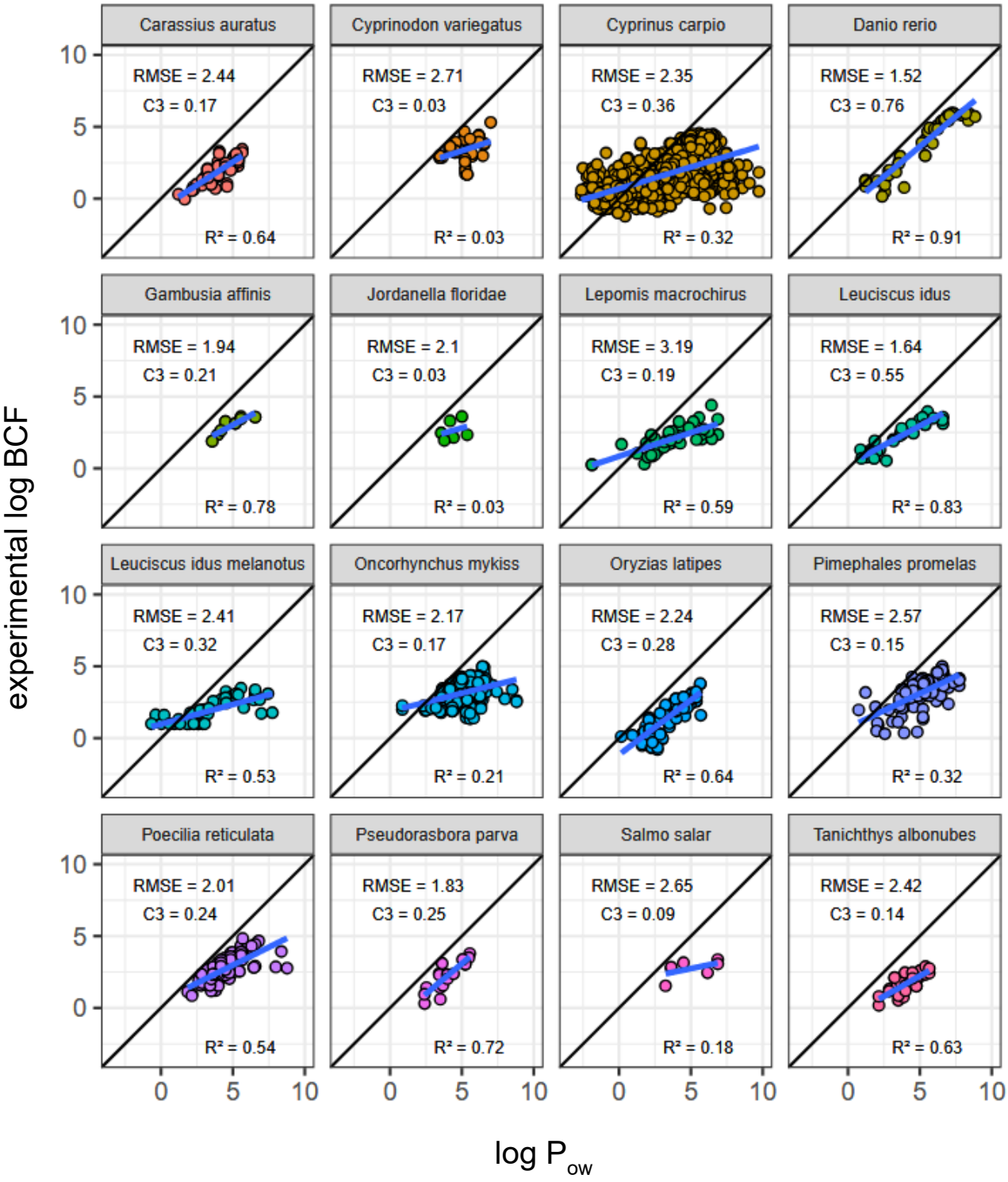


Fig. S12: Experimentally generated log BCF data *vs.* log P_{ow} for 16 fish species. Own compilation of data from public sources, species-specific visualization. Linear regressions (blue lines). RMSE: root mean square error. C3: concordance correlation coefficient as a measure of how well the data set fits the 1:1 line (black bisectors). R^2: Bravais-Pearson variance.

Fig. S13.

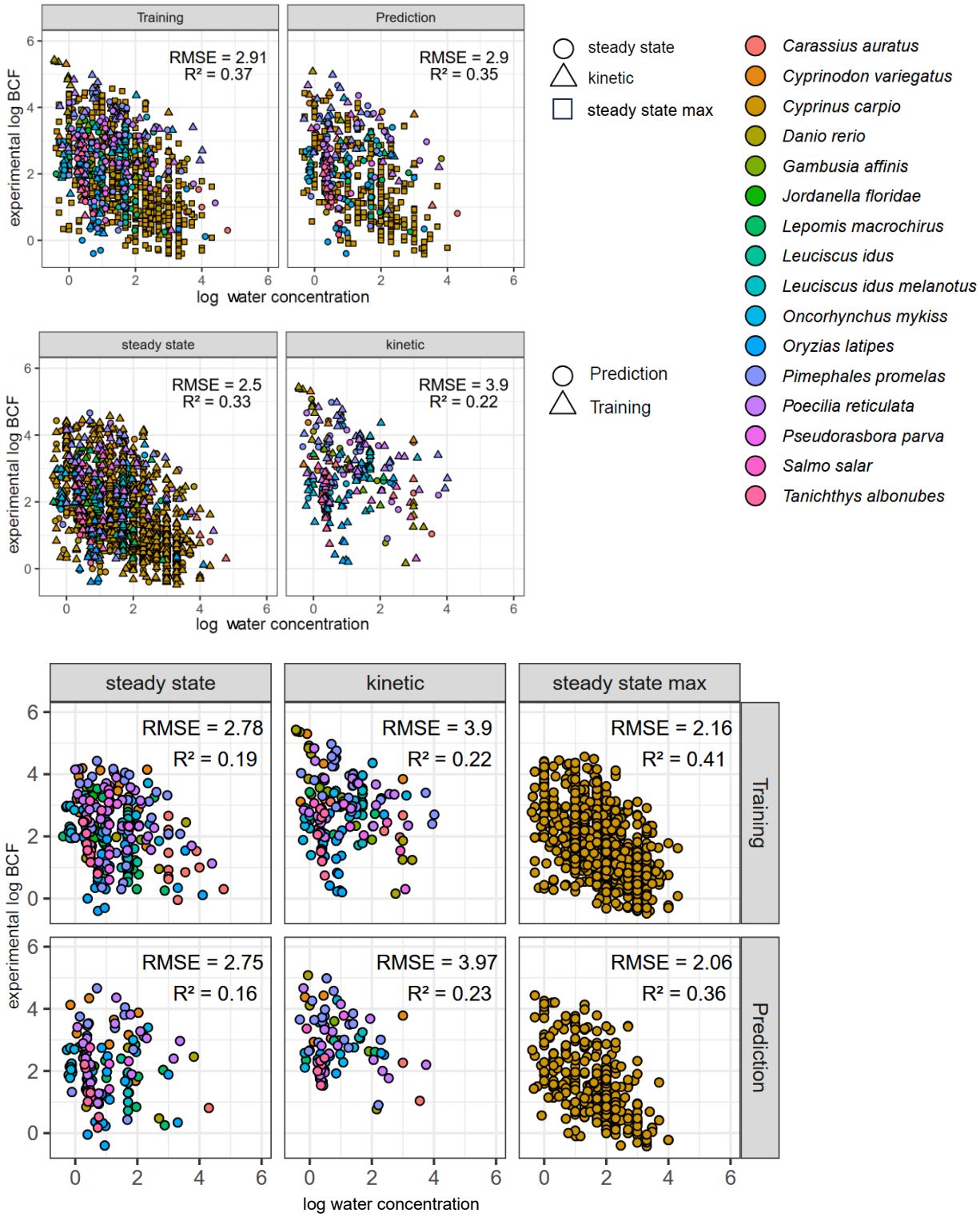


Fig. S13: Experimentally generated log BCF data *vs.* log concentration of corresponding chemicals in the water, own compilation of data from public sources. Separate presentation of data presumably collected at steady state, collected during the kinetics of progressive accumulation or maximum values presumably collected at steady state conditions (circles, triangles and squares, top plots) as well as data used for AI training and those used exclusively for comparison with AI-generated predictions (triangles and circles, middle plots). The bottom four plots show the separation for both criteria (i.e. steady-state and kinetic). The different colors of the symbols refer to the fish species tested. RMSE: root mean square error. R^2: Bravais-Pearson variance. R^2 is in a similar range to that calculated for log Pow *vs.* experimental log BCF, even without considering the lipophilicity of the chemicals (see Fig. S10).

Fig. S14.

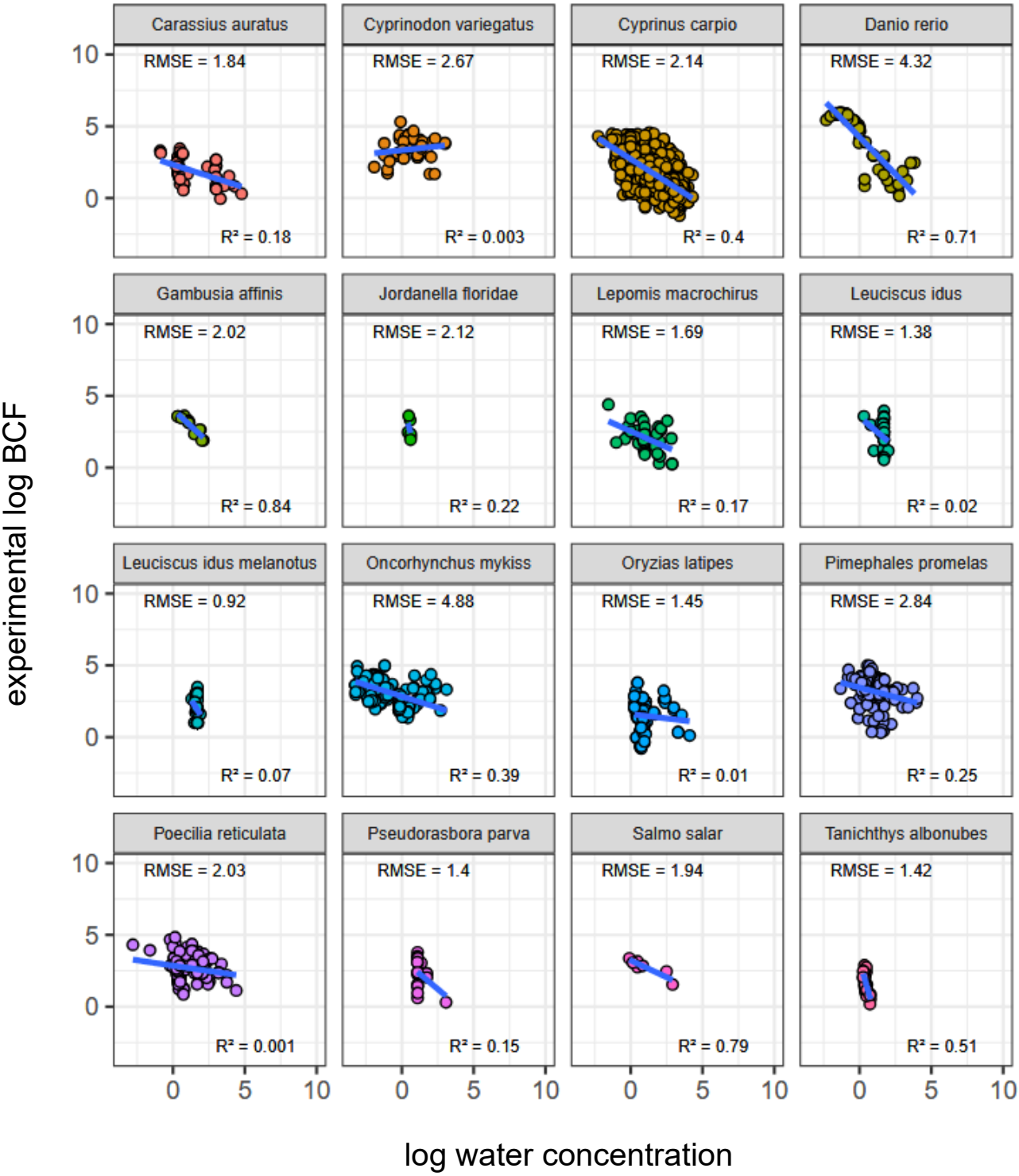


Fig. S14: Experimentally generated log BCF data *vs.* log concentration of corresponding chemicals in the water for 16 fish species. Own compilation of data from public sources, species-specific visualization. Linear regressions (blue lines). RMSE: root mean square error. R^2: Bravais-Pearson variance.

Fig. S15.

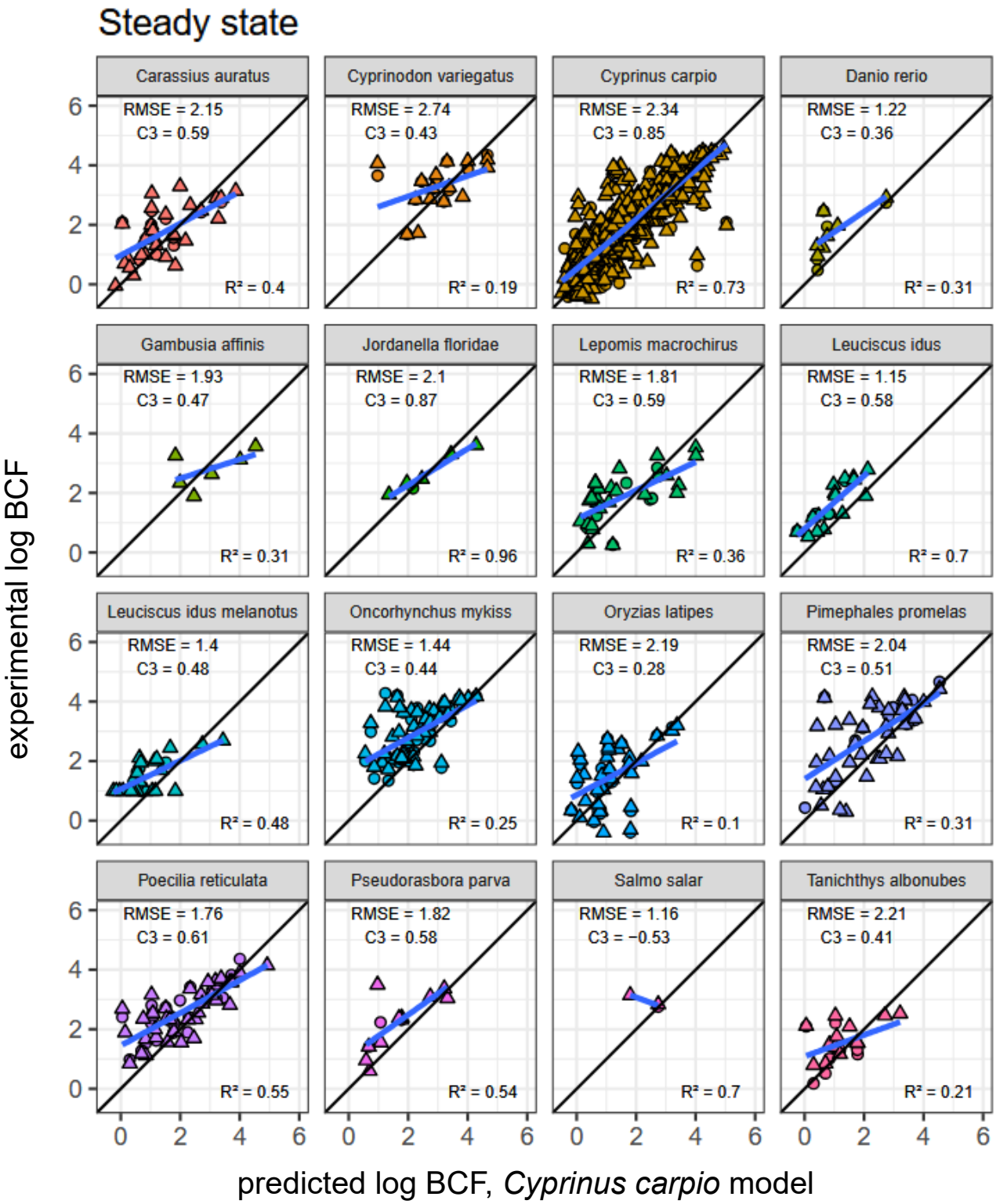

Steady state
Carassius auratus
RMSE = 2.15
C3 = 0.59
R² = 0.4
Cyprinodon variegatus
RMSE = 2.74
C3 = 0.43
R² = 0.19
Cyprinus carpio
RMSE = 2.34
C3 = 0.85
R² = 0.73
Danio rerio
RMSE = 1.22
C3 = 0.36
R² = 0.31
Gambusia affinis
RMSE = 1.93
C3 = 0.47
R² = 0.31
Jordanella floridae
RMSE = 2.1
C3 = 0.87
R² = 0.96
Lepomis macrochirus
RMSE = 1.81
C3 = 0.59
R² = 0.36
Leuciscus idus
RMSE = 1.15
C3 = 0.58
R² = 0.7
Leuciscus idus melanotus
RMSE = 1.4
C3 = 0.48
R² = 0.48
Oncorhynchus mykiss
RMSE = 1.44
C3 = 0.44
R² = 0.25
Oryzias latipes
RMSE = 2.19
C3 = 0.28
R² = 0.1
Pimephales promelas
RMSE = 2.04
C3 = 0.51
R² = 0.31
Poecilia reticulata
RMSE = 1.76
C3 = 0.61
R² = 0.55
Pseudorasbora parva
RMSE = 1.82
C3 = 0.58
R² = 0.54
Salmo salar
RMSE = 1.16
C3 = −0.53
R² = 0.7
Tanichthys albonubes
RMSE = 2.21
C3 = 0.41
R² = 0.21
0
2
4
6
experimental log BCF
predicted log BCF, *Cyprinus carpio* model

Fig. S15 (continued).

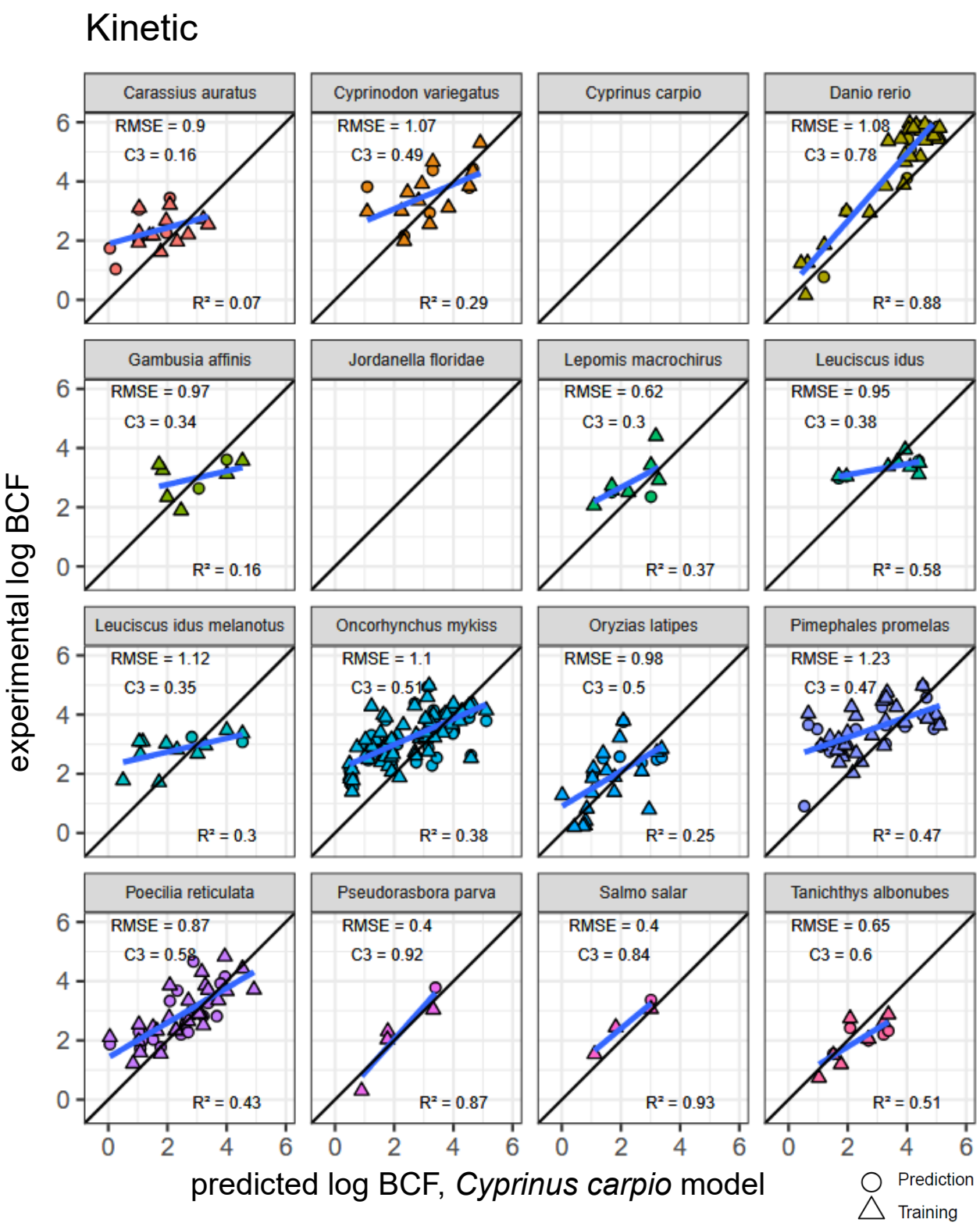


Fig. S15: Experimentally generated log BCF data *vs.* log BCF predictions for 16 fish species made by the DL model that was trained only with data from *Cyprinus carpio*. The varying quality of predictions illustrates the need to integrate species identity into XGBoost and BCFpro for more accurate predictions. Own compilation of data from public sources, species-specific visualization. Experimental data used for AI training (triangles) or used exclusively for comparison with AI-generated predictions (circles). Linear regressions (blue lines). RMSE: root mean square error. C3: concordance correlation coefficient as a measure of how well the data set fits the 1:1 line (black bisectors). R^2: Bravais-Pearson variance. In three species (*Cyprinus carpio*, *Jordanella floridae*, *Salmo salar*) data of some subsets are very limited in number. It should be noted that experimental log BCF data > 5 were only available for *D. rerio*.

Fig. S16.

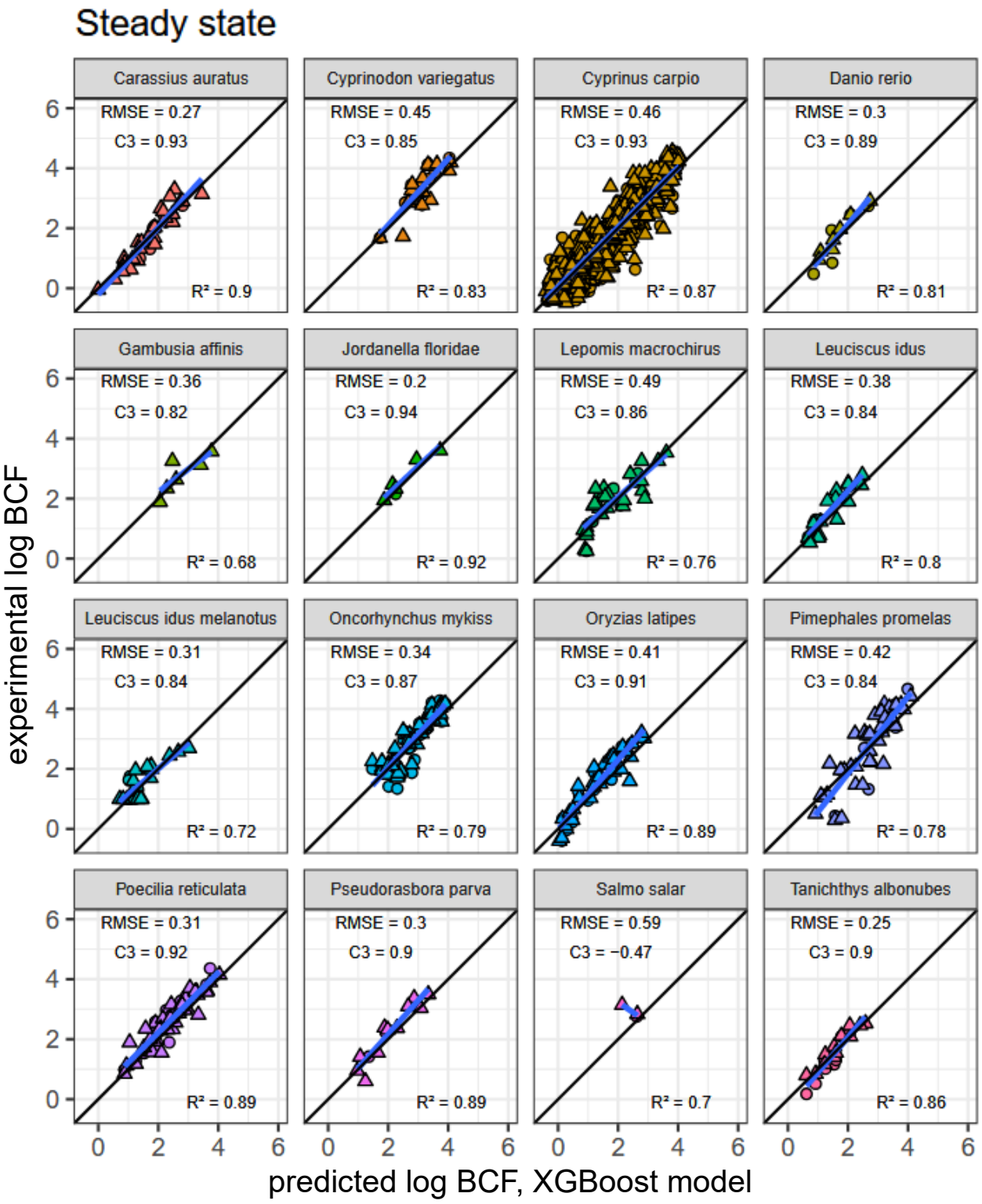

Steady state
Carassius auratus
RMSE = 0.27
C3 = 0.93
R² = 0.9
Cyprinodon variegatus
RMSE = 0.45
C3 = 0.85
R² = 0.83
Cyprinus carpio
RMSE = 0.46
C3 = 0.93
R² = 0.87
Danio rerio
RMSE = 0.3
C3 = 0.89
R² = 0.81
Gambusia affinis
RMSE = 0.36
C3 = 0.82
R² = 0.68
Jordanella floridae
RMSE = 0.2
C3 = 0.94
R² = 0.92
Lepomis macrochirus
RMSE = 0.49
C3 = 0.86
R² = 0.76
Leuciscus idus
RMSE = 0.38
C3 = 0.84
R² = 0.8
Leuciscus idus melanotus
RMSE = 0.31
C3 = 0.84
R² = 0.72
Oncorhynchus mykiss
RMSE = 0.34
C3 = 0.87
R² = 0.79
Oryzias latipes
RMSE = 0.41
C3 = 0.91
R² = 0.89
Pimephales promelas
RMSE = 0.42
C3 = 0.84
R² = 0.78
Poecilia reticulata
RMSE = 0.31
C3 = 0.92
R² = 0.89
Pseudorasbora parva
RMSE = 0.3
C3 = 0.9
R² = 0.89
Salmo salar
RMSE = 0.59
C3 = −0.47
R² = 0.7
Tanichthys albonubes
RMSE = 0.25
C3 = 0.9
R² = 0.86
experimental log BCF
predicted log BCF, XGBoost model
0
2
4
6

Fig. S16 (continued).

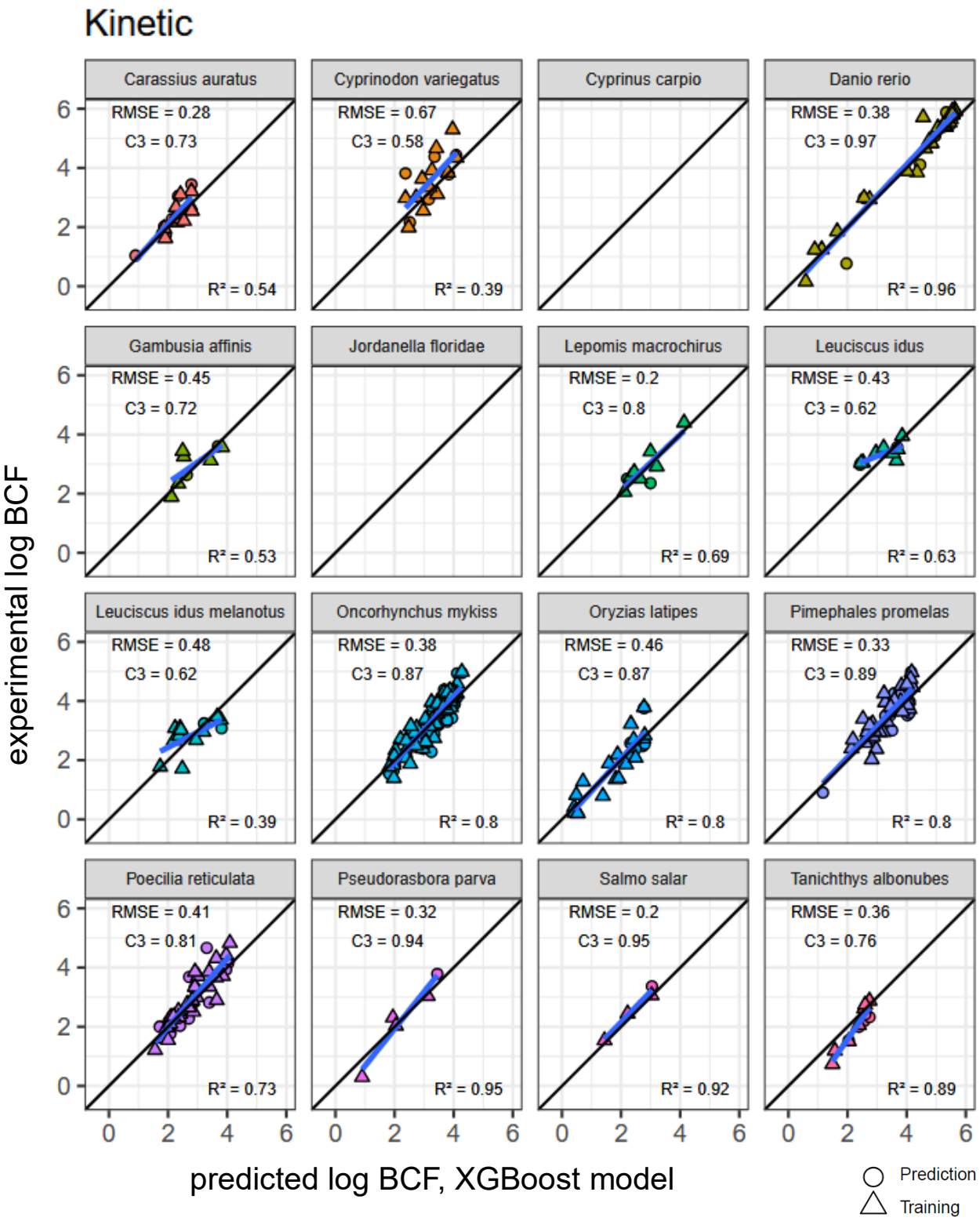


Fig. S16: Experimentally generated log BCF data *vs.* log BCF predictions for 16 fish species made by the XGBoost model. It is worth noting the striking improvements in the predictions, represented by C3, compared to the base model (Figure S13). Own compilation of data from public sources, species-specific visualization. Experimental data used for AI training (triangles) or used exclusively for comparison with AI-generated predictions (circles). Linear regressions (blue lines). RMSE: root mean square error. C3: concordance correlation coefficient as a measure of how well the data set fits the 1:1 line (black bisectors). R^2: Bravais-Pearson variance. Please note that in three species (i.e. *Cyprinus carpio*, *Jordanella floridae, Salmo salar*) data of some subsets are very limited in number.

Fig. S17.

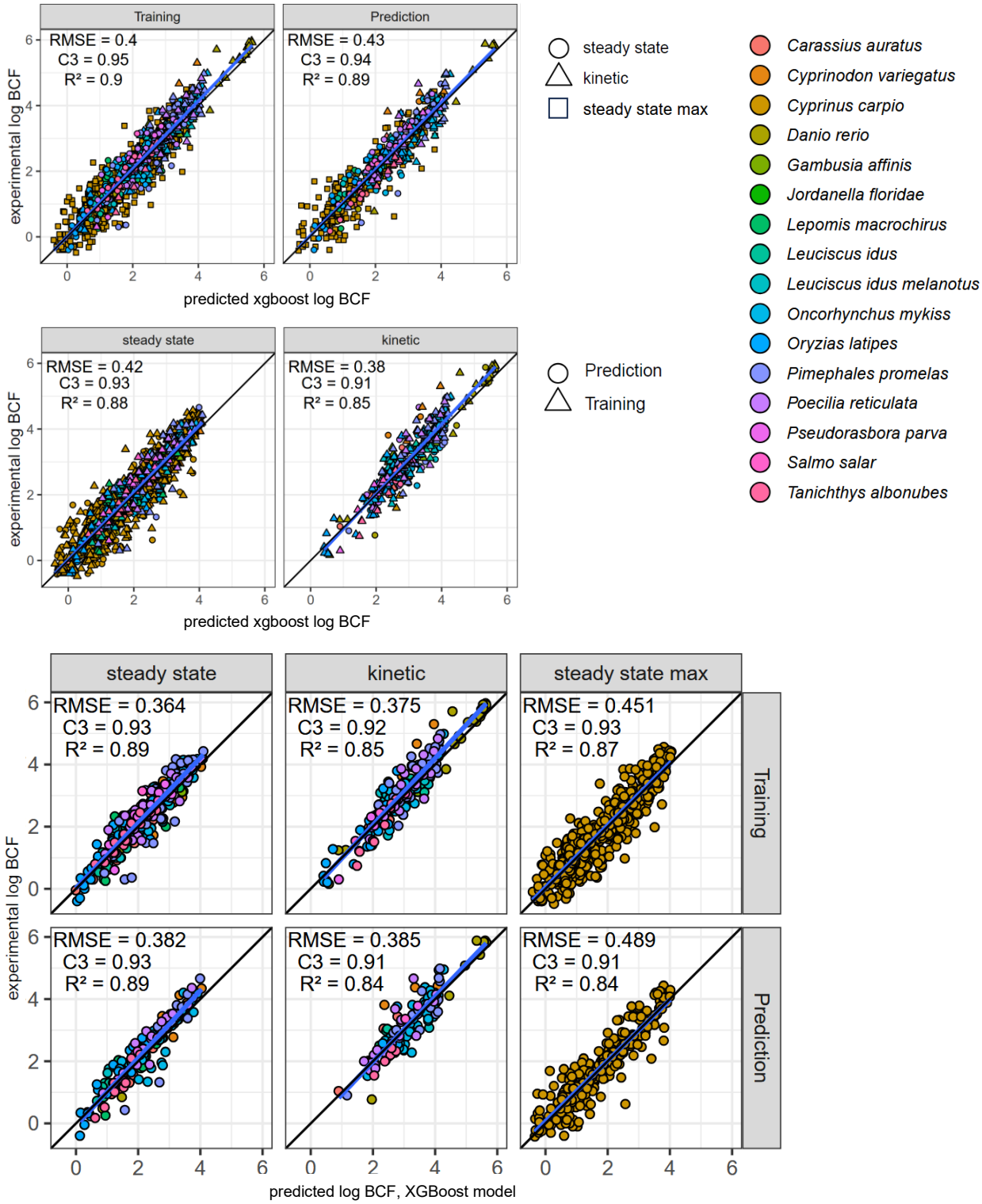


Fig. S17: Experimentally generated log BCF data *vs.* xgboost log BCF predictions of corresponding chemicals generated by BCFpro visualizing the high precision of predictions. Own compilation of data from public sources. Separate presentation of data presumably collected at steady state, collected during the kinetics of progressive accumulation or maximum values presumably collected at steady state conditions (circles, triangles and squares, top plots) as well as data used for AI training and those used exclusively for comparison with AI-generated predictions (triangles and circles, middle plots). The bottom six plots show the separation for both criteria (i.e. steady-state and kinetic). The different colors of the symbols refer to the fish species tested. RMSE: root mean square error. C3: concordance correlation coefficient as a measure of how well the data set fits the 1:1 line (black bisectors). R^2: Bravais-Pearson variance. The two plots in the top row correspond to panels A and B of Fig. 3 in the main paper text.

Fig. S18.

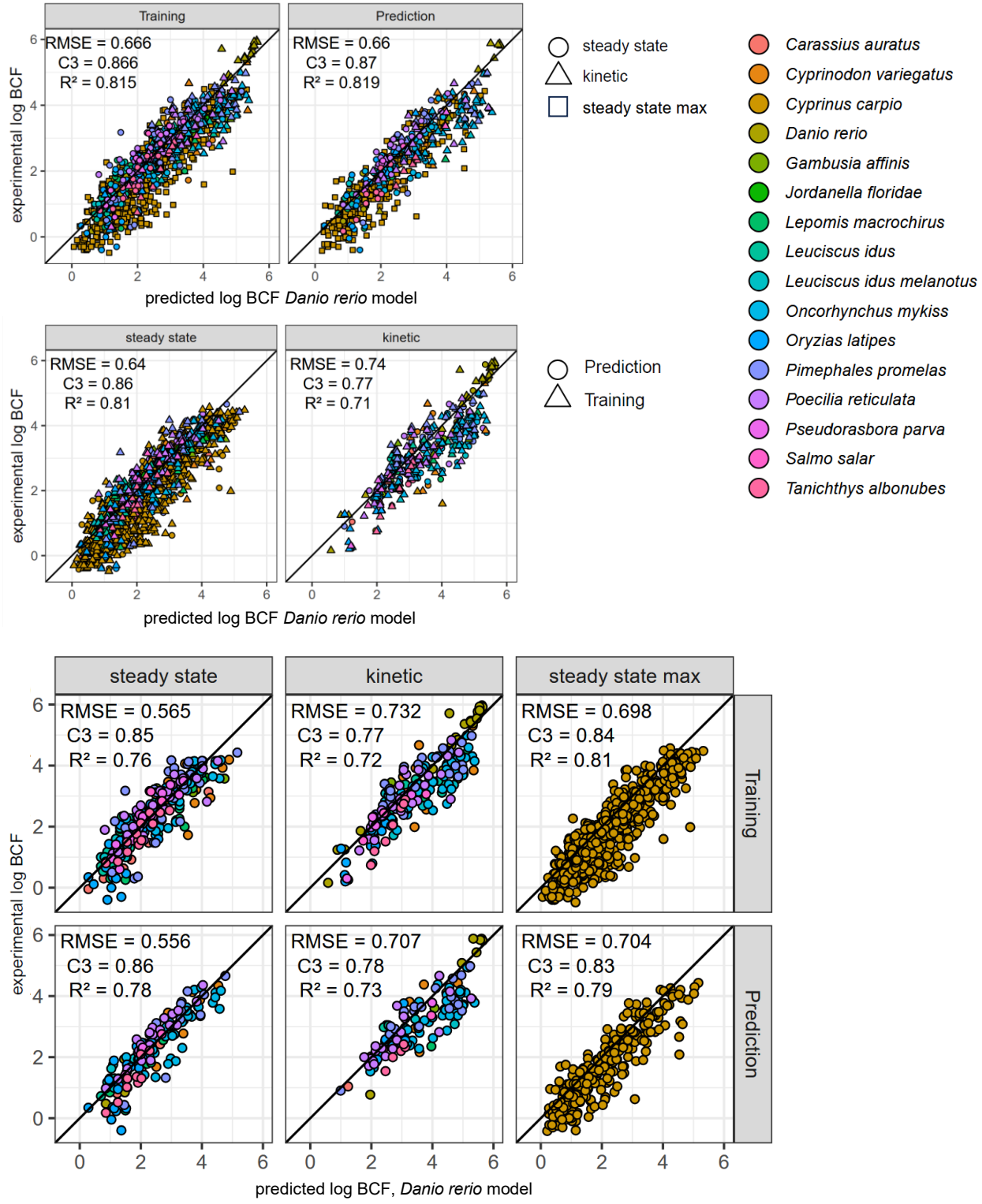


Fig. S18: Experimental log BCF data *vs.* log BCF predictions of corresponding chemicals generated by the XGBoost model optimized for the species *Danio rerio*. Own compilation of data from public sources. Separate presentation of data presumably collected at steady state, collected during the kinetics of progressive accumulation or maximum values presumably collected at steady state conditions (circles, triangles and squares, top plots) as well as data used for AI training and those used exclusively for comparison with AI-generated predictions (triangles and circles, middle plots). The bottom six plots show the separation for both criteria (i.e. steady-state and kinetic). The different colors of the symbols refer to the fish species tested. RMSE: root mean square error. C3: concordance correlation coefficient as a measure of how well the data set fits the 1:1 line (black bisectors). R^2: Bravais-Pearson variance.

Fig. S19.

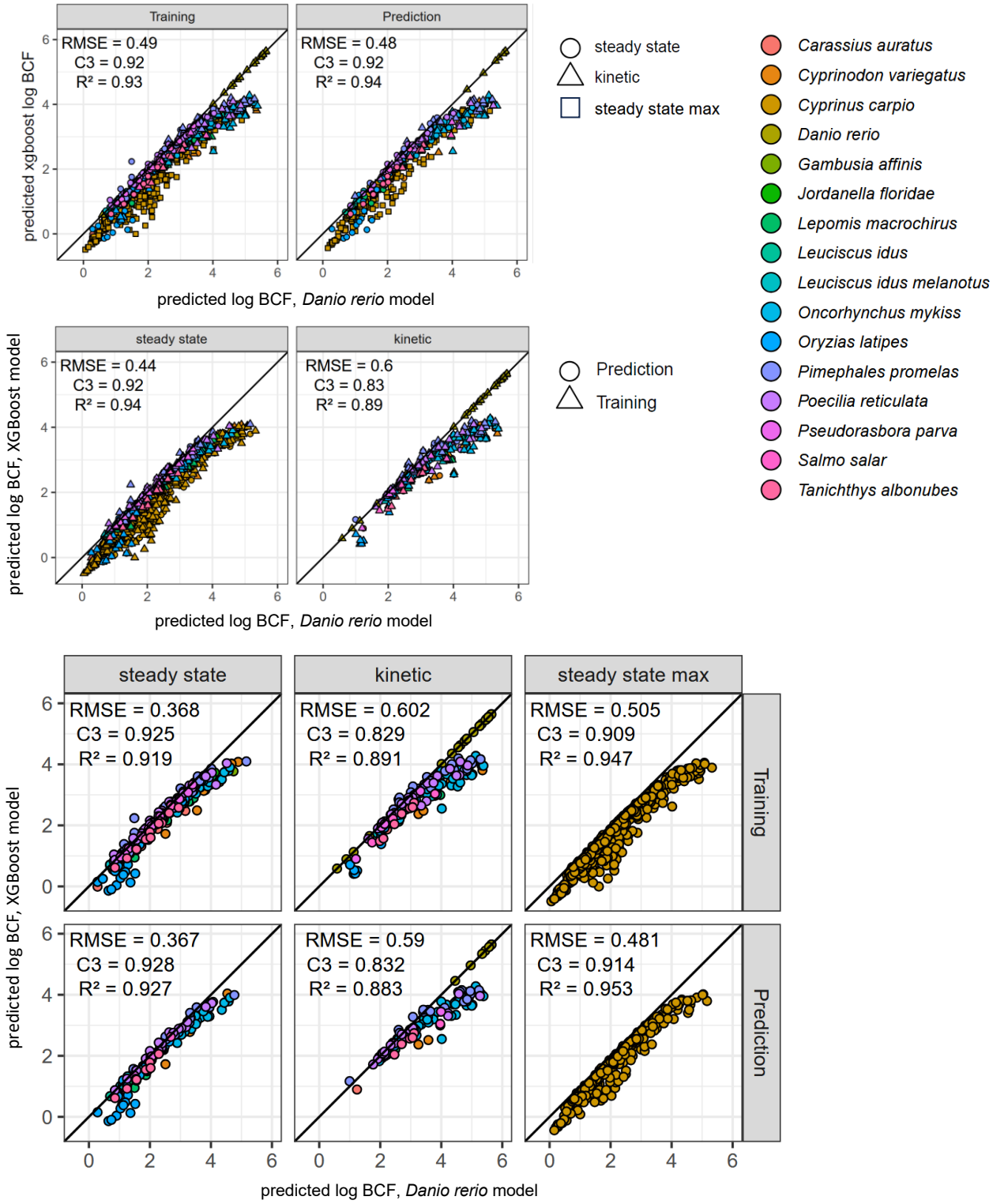


Fig. S19: XGBoost log BCF predictions for specific fish species and test conditions *vs.* log BCF predictions for corresponding chemicals generated by the model optimized for the species *Danio rerio*. Own compilation of data from public sources. Presentation of data in the plots and labels as in Fig. S17. As log BCF data >5 were only available for *Danio rerio* in the training dataset, the *D. rerio* model only reliably reproduces the predictions of the overall XGBoost model calculated for *D. rerio*. In contrast, the *D. rerio* model overpredicts the log BCF values predicted by XGBoost for other fish species in these orders of magnitude, and thus also would overpredict experimentally generated log BCF values for other species, if applied to them. The XGBoost model is clearly superior here.

Fig. S20.

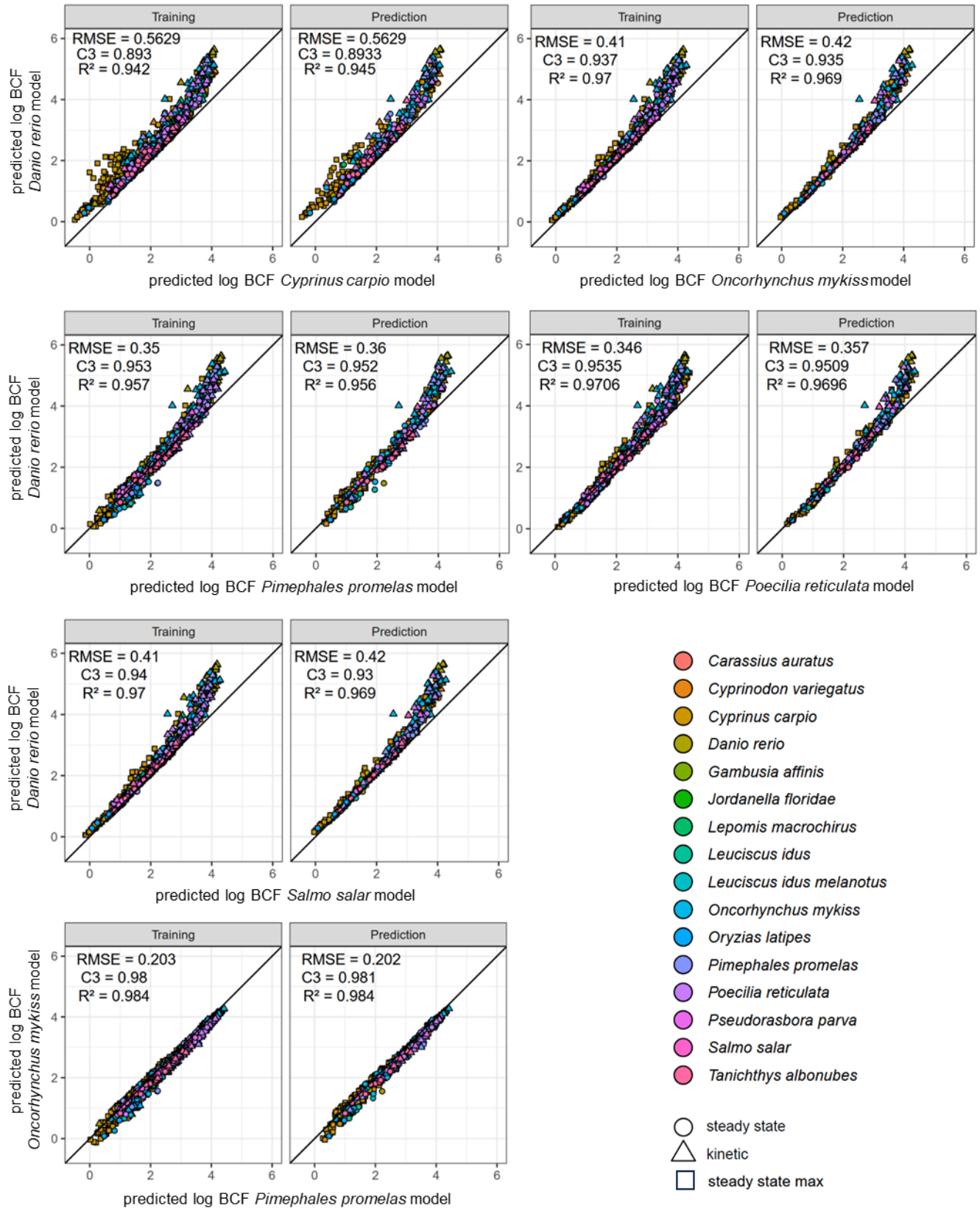


Fig. S20: Correlations of log BCF predictions generated by XGBoost models optimized for different fish species. Own compilation of data from public sources. Separate presentation of data presumably collected at steady state, collected during the kinetics of progressive accumulation or maximum values presumably collected at steady state conditions (circles, triangles and squares). RMSE: root mean square error. C3: concordance correlation coefficient as a measure of how well the data set fits the 1:1 line (black bisectors). R^2: Bravais-Pearson variance. The model optimized for *Danio rerio* overpredicts high log BCF values >4 generated by models optimized for other fish species. This effect does not occur when comparing species-specific models that were not trained with BCF values >4 (exemplarily shown in the bottom plot).

Fig. S21.

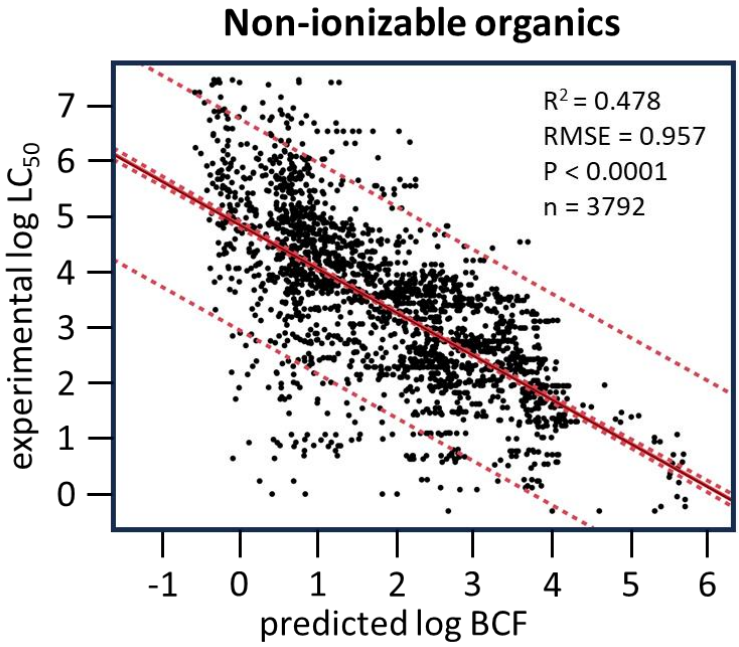


Fig. S21: BCFpro-generated log BCF data *vs.* experimental log LC_{50} of corresponding chemicals, own compilation of data from public sources. Linear regression line (red, solid) and its 95% confidence interval (red dashed, inner interval) plus 95% confidence interval for the data point distribution (red dashed, outer interval). Highly significant (P<0.0001) negative correlation (ANOVA) of the two parameters. R^2: Bravais-Pearson variance. RMSE: root mean square error. n: number of data points. In comparison to Fig. S1, the R^2 improved from 39.4% to 47.8% when predicted log BCF data were used for correlation analysis instead of experimental data.

Table S1. Screening and definitive threshold values for bioaccumulation as applied in several international regulations based on (*1*, *49*, *50*). B= bioaccumulative; vB= very bioaccumulative, Pot B/vB= potentially bioaccumulative or very bioaccumulative.

		Screening criterion (log Kow)		Definitive assessment (BCF values)	
Regulation	**Assessment type**	**Pot B/vB**	**In vivo test required**	**B**	**vB**
POPs UNEP Stockholm Convention	POP identification	5.0	–	5000	–
OSPAR Convention	PBT substances identification	4.0	–	500	–
Canadian Environmental Protection Act (CEPA)	PBT substances identification	5.0	5 .0	5000	–
US Toxic Substances Control Act (TSCA)	PBT/vPvB substances identification	5.0	5.0	1000	5000
Australian National Industrial Chemicals Notification and Assessment Scheme (NICNAS)	PBT/vPvB substances identification	4.2	4.2	2000	5000
China - Measures for the Environmental Management Registration of New Chemical Substances Order n°12, (MEE, 2021)	PBT/vPvB substances identification	3.0	3.0	–	–
Japan Chemical Substances Control Law (CSCL, 2018)	PBT substances identification	3.5	3.5	1000	5000
EU REACH regulation (1907/2006)	PBT/vPvB substances identification	4.5	3.0	2000	5000
EU Plant Protection Product Regulation (1107/2009)	PBT/vPvB substances identification	3.0	3.0	2000	5000
EU Biocidal Products Regulation (528/2012)	PBT/vPvB substances identification	4.5	3.0	2000	5000
EU Human Drug Regulation (EMEA/CHMP/SWP/4447/00 Rev. 1- Corr.*, 08/2024)	PBT/vPvB substances identification	> 4.5	≥ 3.0	2000	5000
EU Veterinary Drug regulation (CVMP/VICH/790/03-FINAL, 10/2004)	PBT/vPvB substances identification	> 4.5	≥ 3.0	2000	5000
UN Globally Harmonised System (GHS)	Hazard classification and labelling	4.0	–	500	–
EU CLP Regulation (1272/2008)	Hazard classification and labelling	4.0	–	500	–

Table S2. Existing models for in silico generation of BCF data.

Name (Authors)	No of BCF entries	No of species	Type of Analysis	R^2 (training / test datasets)	RMSE (training / test datasets)	Reference
R code in Github (Zang et al.)	608	unknown	Support vector regression (SVR)	0.974 / 0.879	0.188 / 0.418	(*31*)
CORAL based (Toropova et al.)	237	unknown	Monte Carlo optimization - MLR	0.901 / 0.916	0.446 / 0.440	(*32*)
CORAL based (Toropov et al.)	1035	unknown	Monte Carlo optimization - MLR	0.775 / 0.828	0.627 / 0.545	(*33*)
- (Lu et al.)	214	unknown	Polynomial regression	0.810 / -	0.614 / -	(*48*)
- (Sacan et al.)	122	unknown	MLR	0.926 / -	0.584 / -	(*49*)
- (Arnot & Quinn)	869	primary 2	Linear regression	> 0.7 / -	- / -	(*50*)
- (Ivanciuc et al.)	58	4	Polynomial spline	0.959 / -	0.245 / -	(*51*)
- (Pramanik & Roy)	522	unknown	MLR / partial least squares regression	0.614 / 0.696	- / 0.738	(*34*)
CAESAR (Zhao et al.)	473	unknown	Radial basis function neural network	0.830 / 0.800	0.560 / 0.590	(*35*)
BAFBCF/EPISuite (Weisbrod et al.)	527	unknown	Regression analysis	0.800 / 0.770	0.620 / 0.610	(*17*)
OPERA (Mansouri et al.)	626	unknown	K-nearest neighbor (kNN)	0.850 / 0.830	0.530 / 0.640	(*36*)
- (Lunghini et al.)	1129	unknown	SVR	0.750 / 0,770	0.680 / 0.760	(*37*)
- (Dimitrov et al.)	443	unknown	Polynomial regression	0.730 / -	0.650 / -	(*52*)
- (Piir et al.)	473	unknown	MLR	0.751 / 0.616	0.669 / 0.827	(*38*)
- (Miller et al.)	352	1	4-layer multilayer perceptrons	0.887 / 0.702	0.403 / 0.524	(*39*)
- (Xu et al.)	3240	multiple	4-layer multilayer perceptrons	0.881 / 0.618	0.209 / 0.627	(*42*)
CDDD (Zhao et al.)	1056	unknown	Random forest (RF), SVR, XGBoost	0.703 / 0.676	0.736 / 0.593	(*27*)
- (Zhu et al.)	1724	unknown	RF, SVR, XGBoost	0.996 / 0.753	0.089 / 0.701	(*30*)
- (Kobayashi & Yoshida)	522	unknown	RF, SVR, XGBoost	0.923 / 0.863	0.378 / 0.494	(*53*)
Java-based app (Pore et al.)	575	unknown	MLR, RF, SVR, XGBoost	0.790 / -	0.460 / 0.510	(*54*)
Java-based app (Chandra et al.)	2411	unknown	LSVM, MLR, Commercial descriptors "alvaDesc", Q-RASPR	0.943 / -	- / 0.656	(*55*)
BCFpro app (Aalizadeh in Köhler et al.)	3826	16	Convolutional neural network, XGBoost	0.901 / 0.883	0.402 / 0.436	This study

References Supplementary Material

1. P. N. H. Wassenaar, E. M. J. Verbruggen, E. Cieraad, W. J. G. M. Peijnenburg, M. G. Vijver, Variability in fish bioconcentration factors: Influences of study design and consequences for regulation. *Chemosphere* **239**, 124731 (2020).

2. J. A. Arnot, F. A. P. C. Gobas, A review of bioconcentration factor (BCF) and bioaccumulation factor (BAF) assessments for organic chemicals in aquatic organisms. *Environmental Reviews* **14**, 257–297 (2006).

3. H.-R. Köhler, T. Gräff, M. Schweizer, J. Blumhardt, J. Burkhardt, L. Ehmann, J. Hebel, C. Heid, L. Kundy, J. Kuttler, M. Malusova, F.-M. Moroff, A.-F. Schlösinger, P. Schulze-Berge, E. I. Panagopoulou, D. E. Damalas, N. S. Thomaidis, R. Triebskorn, D. Maletzki, U. Kühnen, P. C. von der Ohe, 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 Research* **235**, 119864 (2023).

4. N. Rodriguez-Sanchez, M. T. D. Cronin, A. Lillicrap, J. C. Madden, P. Piechota, K. E. Tollefsen, Development of a list of reference chemicals for evaluating alternative methods to in vivo fish bioaccumulation tests. *Environmental Toxicology and Chemistry* **33**, 2740–2752 (2014).

5. J. M. Armitage, J. A. Arnot, F. Wania, D. Mackay, Development and evaluation of a mechanistic bioconcentration model for ionogenic organic chemicals in fish. *Environmental Toxicology and Chemistry* **32**, 115–128 (2013).

6. S. M. Lundberg, B. Nair, M. S. Vavilala, M. Horibe, M. J. Eisses, T. Adams, D. E. Liston, D. K.-W. Low, S.-F. Newman, J. Kim, S.-I. Lee, Explainable machine-learning predictions for the prevention of hypoxaemia during surgery. *Nat Biomed Eng* **2**, 749–760 (2018).

7. S. Mirjalili, S. M. Mirjalili, A. Lewis, Grey Wolf Optimizer. *Advances in Engineering Software* **69**, 46–61 (2014).

8. A. J. Siccardi, H. W. Garris, W. T. Jones, D. B. Moseley, L. R. D'Abramo, S. A. Watts, Growth and Survival of Zebrafish (Danio rerio) Fed Different Commercial and Laboratory Diets. *Zebrafish* **6**, 275–280 (2009).

9. C. Schlechtriem, A. Fliedner, C. Schäfers, Determination of lipid content in fish samples from bioaccumulation studies: contributions to the revision of guideline OECD 305. *Environmental Sciences Europe* **24**, 13 (2012).

10. C. U. Johnston, C. J. Kennedy, A Review of P-Glycoprotein Function and Regulation in Fish. *Fishes* **9**, 51 (2024).

11. R. Silva, V. Vilas-Boas, H. Carmo, R. J. Dinis-Oliveira, F. Carvalho, M. de Lourdes Bastos, F. Remião, Modulation of P-glycoprotein efflux pump: induction and activation as a therapeutic strategy. *Pharmacology & Therapeutics* **149**, 1–123 (2015).

12. J. L. M. Hermens, J. H. M. de Bruijn, D. N. Brooke, The octanol–water partition coefficient: Strengths and limitations*. *Environmental Toxicology and Chemistry* **32**, 732–733 (2013).

13. ECHA, Guidance on Information Requirements and Chemical Safety Assessment. Chapter R.11: PBT/vPvB assessment. Version 4.0. (2023).

14. F. Grisoni, V. Consonni, S. Villa, M. Vighi, R. Todeschini, QSAR models for bioconcentration: Is the increase in the complexity justified by more accurate predictions? *Chemosphere* **127**, 171–179 (2015).

15. ECHA, "Guidance on Information Requirements and Chemical Safety Assessment Chapter R.7c: Endpoint specific guidance. Vers. 4.0" (2023); https://echa.europa.eu/documents/10162/17224/information_requirements_r7c_en.pdf.

16. S. Endo, J. Bauerfeind, K.-U. Goss, Partitioning of neutral organic compounds to structural proteins. *Environmental Science and Technology* **46**, 12697–12703 (2012).

17. A. V. Weisbrod, L. P. Burkhard, J. Arnot, O. Mekenyan, P. H. Howard, C. Russom, R. Boethling, Y. Sakuratani, T. Traas, T. Bridges, C. Lutz, M. Bonnell, K. Woodburn, T. Parkerton, Workgroup report: Review of fish bioaccumulation databases used to identify persistent, bioaccumlative, toxic substances. *Environmental Health Perspectives* **115**, 255–261 (2007).

18. O. V. Martin, J. Adams, A. Beasley, S. Belanger, R. L. Breton, T. C. M. Brock, V. A. Buonsante, M. Galay Burgos, J. Green, P. D. Guiney, T. Hall, M. Hanson, M. J. Harris, T. R. Henry, D. Huggett, M. Junghans, R. Laskowski, G. Maack, C. T. A. Moermond, G. Panter, A. Pease, V. Poulsen, M. Roberts, C. Rudén, C. E. Schlekat, I. Schoeters, K. R. Solomon, J. Staveley, B. Stubblefield, J. P. Sumpter, M. S. J. Warne, R. Wentsel, J. R. Wheeler, B. A. Wolff, K. Yamazaki, H. Zahner, M. Ågerstrand, Improving environmental risk assessments of chemicals: Steps towards evidence-based ecotoxicology. *Environment International* **128**, 210–217 (2019).

19. G. Stieger, M. Scheringer, C. A. Ng, K. Hungerbühler, Assessing the persistence, bioaccumulation potential and toxicity of brominated flame retardants: data availability and quality for 36 alternative brominated flame retardants. *Chemosphere* **116**, 118 (2014).

20. J. Glüge, B. I. Escher, M. Scheringer, How error-prone bioaccumulation experiments affect the risk assessment of hydrophobic chemicals and what could be improved. *Integrated Environmental Assessment and Management* **19**, 792–803 (2023).

21. OECD, 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. (2012).

22. L. Bertato, N. Chirico, E. Papa, Predicting the Bioconcentration Factor in Fish from Molecular Structures. *Toxics* **10** (2022).

23. OECD, Guidance Document on Aspects of OECD TG 305 on Fish Bioaccumulation. https://www.oecd-ilibrary.org/content/publication/c25ad843-en (2017). https://www.oecd-ilibrary.org/content/publication/c25ad843-en.

24. L. Böhm, R.-A. Düring, H.-J. Bruckert, C. Schlechtriem, Can solid-phase microextraction replace solvent extraction for water analysis in fish bioconcentration studies with highly hydrophobic organic chemicals? *Environmental Toxicology and Chemistry* **36**, 2887–2894 (2017).

25. F. A. P. C. Gobas, Y.-S. Lee, Growth-Correcting the Bioconcentration Factor and Biomagnification Factor in Bioaccumulation Assessments. *Environmental Toxicology and Chemistry* **38**, 2065–2072 (2019).

26. M. Adolfsson-Erici, G. Åkerman, M. S. Mclachlan, Measuring bioconcentration factors in fish using exposure to multiple chemicals and internal benchmarking to correct for growth dilution. *Environmental Toxicology and Chemistry* **31**, 1853–1860 (2012).

27. L. Zhao, F. Montanari, H. Heberle, S. Schmidt, Modeling bioconcentration factors in fish with explainable deep learning. *Artificial Intelligence in the Life Sciences* **2** (2022).

28. K. von Borries, H. Holmquist, M. Kosnik, K. V. Beckwith, O. Jolliet, J. M. Goodman, P. Fantke, Potential for Machine Learning to Address Data Gaps in Human Toxicity and Ecotoxicity Characterization. *Environ. Sci. Technol.* **57**, 18259–18270 (2023).

29. P. Fantke, J. A. Arnot, W. J. Doucette, Improving plant bioaccumulation science through consistent reporting of experimental data. *Journal of Environmental Management* **181**, 374–384 (2016).

30. M. Zhu, Z. Xiao, T. Zhang, G. Lu, Construction of interpretable ensemble learning models for predicting bioaccumulation parameters of organic chemicals in fish. *Journal of Hazardous Materials* **482**, 136606 (2025).

31. Q. Zang, K. Mansouri, A. J. Williams, R. S. Judson, D. G. Allen, W. M. Casey, N. C. Kleinstreuer, In Silico Prediction of Physicochemical Properties of Environmental Chemicals Using Molecular Fingerprints and Machine Learning. *J. Chem. Inf. Model.* **57**, 36–49 (2017).

32. A. P. Toropova, A. A. Toropov, S. E. Martyanov, E. Benfenati, G. Gini, D. Leszczynska, J. Leszczynski, CORAL: Monte Carlo Method as a Tool for the Prediction of the Bioconcentration Factor of Industrial Pollutants. *Molecular Informatics* **32**, 145–154 (2013).

33. A. A. Toropov, A. P. Toropova, A. Lombardo, A. Roncaglioni, E. Benfenati, G. Gini, CORAL: Building up the model for bioconcentration factor and defining it's applicability domain. *European Journal of Medicinal Chemistry* **46**, 1400–1403 (2011).

34. S. Pramanik, K. Roy, Modeling bioconcentration factor (BCF) using mechanistically interpretable descriptors computed from open source tool "PaDEL-Descriptor." *Environ Sci Pollut Res* **21**, 2955–2965 (2014).

35. C. Zhao, E. Boriani, A. Chana, A. Roncaglioni, E. Benfenati, A new hybrid system of QSAR models for predicting bioconcentration factors (BCF). *Chemosphere* **73**, 1701–1707 (2008).

36. K. Mansouri, C. M. Grulke, R. S. Judson, A. J. Williams, OPERA models for predicting physicochemical properties and environmental fate endpoints. *Journal of Cheminformatics* **10** (2018).

37. F. Lunghini, G. Marcou, P. Azam, R. Patoux, M. H. Enrici, F. Bonachera, D. Horvath, A. Varnek, QSPR models for bioconcentration factor (BCF): are they able to predict data of industrial interest? *SAR and QSAR in Environmental Research* **30**, 507–524 (2019).

38. G. Piir, S. Sild, A. Roncaglioni, E. Benfenati, U. Maran, QSAR model for the prediction of bio-concentration factor using aqueous solubility and descriptors considering various electronic effects. *SAR and QSAR in Environmental Research* **21**, 711–729 (2010).

39. T. H. Miller, M. D. Gallidabino, J. R. MacRae, S. F. Owen, N. R. Bury, L. P. Barron, Prediction of bioconcentration factors in fish and invertebrates using machine learning. *Science of the Total Environment* **648**, 80–89 (2019).

40. M. T. Saçan, S. S. Erdem, G. A. Özpınar, I. A. Balcıoglu, 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 (2004).

41. S. D. Dimitrov, N. C. Dimitrova, J. D. Walker, G. D. Veith, O. G. Mekenyan, Predicting bioconcentration factors of highly hydrophobic chemicals. Effects of molecular size. *Pure and Applied Chemistry* **74**, 1823–1830 (2002).

42. J.-Y. Xu, K. Wang, S.-H. Men, Y. Yang, Q. Zhou, Z.-G. Yan, QSAR-QSIIR-based prediction of bioconcentration factor using machine learning and preliminary application. *Environment International* **177**, 108003 (2023).

43. R. Gozalbes, J. P. Doucet, F. Derouin, Application of Topological Descriptors in QSAR and Drug Design: History and New Trends. *Current Drug Targets - Infectious Disorders* **2**, 93–102 (2002).

44. P. Ball, Is AI leading to a reproducibility crisis in science? *Nature* **624**, 22–25 (2023).

45. R. Aalizadeh, V. Nikolopoulou, N. S. Thomaidis, Development of Liquid Chromatographic Retention Index Based on Cocamide Diethanolamine Homologous Series (C(n)-DEA). *Anal. Chem.* **94**, 15987–15996 (2022).

46. S. Gimeno, D. Allan, P. Kai, P. Remuzat, M. Collard, Are current regulatory log Kow cut-off values fit-for-purpose as a screening tool for bioaccumulation potential in aquatic organisms? *Regulatory Toxicology and Pharmacology* **147** (2024).

47. J. R. Centre, I. for H. and C. Protection, S. Gottardo, N. Hartmann, B. Sokull-Klüttgen, Review of available criteria for non-aquatic organisms within PBT/vPvB frameworks – Part II: toxicity assessment, Publications Office, 2014, https://data.europa.eu/doi/10.2788/99200. (2014).

48. X. Lu, S. Tao, H. Hu, R. W. Dawson, Estimation of bioconcentration factors of nonionic organic compounds in fish by molecular connectivity indices and polarity correction factors. *Chemosphere* **41**, 1675–1688 (2000).

49. M. T. Saçan, S. S. Erdem, G. A. Özpınar, I. A. Balcıoglu, 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 (2004).

50. J. A. Arnot, C. L. Quinn, Development and Evaluation of a Database of Dietary Bioaccumulation Test Data for Organic Chemicals in Fish. *Environ. Sci. Technol.* **49**, 4783–4796 (2015).

51. T. Ivanciuc, O. Ivanciuc, D. J. Klein, Modeling the bioconcentration factors and bioaccumulation factors of polychlorinated biphenyls with posetic quantitative super-structure/activity relationships (QSSAR). *Mol Divers* **10**, 133–145 (2006).

52. S. Dimitrov, R. Breton†, D. MacDonald†, J. D. Walker, O. Mekenyan, Quantitative prediction of biodegradability, metabolite distribution and toxicity of stable metabolites. *SAR and QSAR in Environmental Research* **13**, 445–455 (2002).

53. Y. Kobayashi, K. Yoshida, Development of QSAR models for prediction of fish bioconcentration factors using physicochemical properties and molecular descriptors with machine learning algorithms. *Ecological Informatics* **63**, 101285 (2021).

54. S. Pore, A. Pelloux, M. Chatterjee, A. Banerjee, K. Roy, 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. *Journal of Hazardous Materials* **479**, 135725 (2024).

55. A. Chandra, A. Banerjee, K. Roy, Prediction of bioconcentration factors (BCFs) and bioaccumulation factors (BAFs) for per- and polyfluoroalkyl substances (PFASs) using Read-Across and q-RASPR. *Science of The Total Environment* **993**, 180007 (2025).