TEXTE
117/2026

Final report

Integrated strategy for the assessment of aquatic and terrestrial bioaccumulation

by:

Christian Schlechtriem
Fraunhofer Institute for Molecular Biology and Applied Ecology IME, Schmallenberg

Monika Nendza
Analytisches Laboratorium für Umweltuntersuchungen und Auftragsforschung (AL-Luhnstedt), Luhnstedt

Verena Kosfeld
Fraunhofer Institute for Molecular Biology and Applied Ecology IME, Schmallenberg

publisher:
German Environment Agency

German Environment Agency

Umwelt Bundesamt

TEXTE 117/2026

REFOPLAN of the Federal Ministry for the Environment, Nature Conservation, Nuclear Safety and Consumer Protection

Project No. (FKZ) 3722 65 402 0

Final report

Integrated strategy for the assessment of aquatic and terrestrial bioaccumulation

by

Christian Schlechtriem
Fraunhofer Institute for Molecular Biology and Applied Ecology IME, Schmallenberg

Monika Nendza
Analytisches Laboratorium für Umweltuntersuchungen und Auftragsforschung (AL-Luhnstedt), Luhnstedt

Verena Kosfeld
Fraunhofer Institute for Molecular Biology and Applied Ecology IME, Schmallenberg

On behalf of the German Environment Agency

Imprint

Publisher
Umweltbundesamt
Wörlitzer Platz 1
06844 Dessau-Roßlau
Tel: +49 340-2103-0
Fax: +49 340-2103-2285
buergerservice@uba.de
Internet: www.umweltbundesamt.de

Report performed by:
Fraunhofer Institute for Molecular Biology and Applied Ecology IME
Auf dem Aberg 1
57392 Schmallenberg
Germany

Report completed in:
October 2025

Edited by:
Section IV Chemical Safety – Coordination Group bioaccumulation and ecotoxicology
Caren Rauert & Daniel Frein

DOI:
https://doi.org/10.60810/openumwelt-8221

ISSN 1862-4804

Dessau-Roßlau, July 2026

Abstract: Integrated strategy for the assessment of aquatic and terrestrial bioaccumulation

The evaluation of bioaccumulation potential is one of the fundamental elements of environmental risk assessment. The report proposes a comprehensive framework for modernising the assessment of bioaccumulation potential in chemicals under European regulatory systems. It addresses the limitations of traditional fish bioconcentration testing (OECD TG 305) and proposes an integrated, tiered strategy combining in silico, in vitro, and invertebrate approaches to minimise vertebrate testing while maintaining the level of environmental protection. The project, funded by the German Environment Agency (UBA), evaluated 26 representative substances across diverse chemical classes using fish, amphipod (*Hyalella azteca*, OECD TG 321), and in vitro hepatic biotransformation assays (OECD TG 319A/B), supported by QSAR modelling and physicochemical parameter analysis (log K_{OW}, log K_{OA}, p*K*a) for comparability of test results. The integrated assessment scheme emphasises early screening, use of in silico data and further New Approach Methodologies (NAMs), to ensure minimal animal testing. The study shows that NAMs can effectively exclude bioaccumulation concerns under certain conditions, reducing reliance on animal tests. Yet, for challenging substances, critical review and complementary methods remain essential. Special guidance is provided for complex substances, including ionisable, amphiphilic, and superhydrophobic compounds, and for assessing accumulation in air-breathing organisms. Monitoring data and additional information from further studies should also be considered, if available, as part of a Weight of Evidence (WoE) approach. The project demonstrates that integrated, evidence-based bioaccumulation assessment is feasible, robust, and aligned with the 3R principles, marking a decisive step towards animal-free chemical regulation in Europe. Concluding reflections emphasise the scientific, ethical, and regulatory value of NAMs and the possibility of harmonised criteria and assessment strategies across EU chemicals regulations.

Kurzbeschreibung: Integrated strategy for the assessment of aquatic and terrestrial bioaccumulation

Die Bewertung des Bioakkumulationspotenzials ist eines der grundlegenden Elemente der Umweltrisikobewertung. Der Bericht schlägt ein Konzept für die Modernisierung der Bewertung des Bioakkumulationspotenzials von Chemikalien im Rahmen der europäischen Regulierungssysteme vor. Er befasst sich mit den Grenzen der traditionellen Biokonzentrationsprüfung an Fischen (OECD TG 305) und beschreibt eine integrierte, mehrstufige Strategie, die In-silico-, In-vitro- und Wirbellosen-Ansätze kombiniert, um die Prüfung an Wirbeltieren zu minimieren und gleichzeitig das derzeitige Umweltschutzniveau aufrechtzuerhalten. Das vom Umweltbundesamt (UBA) finanzierte Projekt bewertete 26 repräsentative Substanzen aus verschiedenen chemischen Klassen unter Verwendung von Fischen, Amphipoden (*Hyalella azteca*, OECD TG 321) und In-vitro-Leberbiotransformationsassays (OECD TG 319A/B), unterstützt durch QSAR-Modellierung und physikalisch-chemische Parameteranalyse (log K_{OW}, log K_{OA}, p*K*a) in Bezug auf die Vergleichbarkeit der Ergebnisse. Das integrierte Bewertungssystem legt den Schwerpunkt auf Screening-Verfahren, die Verwendung von In-silico-Daten und weiteren neuen Methoden (New Approach Methodologies, NAMs), um Tierversuche auf ein Minimum zu reduzieren. Die Studie zeigt, dass NAMs unter bestimmten Bedingungen Bedenken hinsichtlich der Bioakkumulation von Stoffen wirksam ausschließen können, wodurch die Abhängigkeit von Tierversuchen verringert wird. Bei problematischen Stoffen sind jedoch weiterhin kritische Überprüfungen und ergänzende Methoden unerlässlich. Spezielle Leitlinien werden für komplexe Stoffe, darunter ionisierbare, amphiphile und superhydrophobe Verbindungen, sowie für die Bewertung der Akkumulation in luftatmenden Organismen bereitgestellt. Monitoringdaten und zusätzliche Informationen aus weiteren Studien werden ebenfalls, sofern verfügbar, im Rahmen eines WoE-Ansatzes berücksichtigt. Das Projekt zeigt, dass eine integrierte, evidenzbasierte Bewertung der Bioakkumulation machbar, robust und mit den 3R-Prinzipien vereinbar ist, was einen entscheidenden Schritt in Richtung einer tierversuchsfreien Chemikalienregulierung in Europa darstellt. Abschließende Überlegungen betonen den wissenschaftlichen, ethischen und regulatorischen Wert der NAMs und die Möglichkeit harmonisierter Kriterien und Bewertungsstrategien in allen EU-Chemikalienregulierungen.

Table of content

List of figures

List of tables

List of abbreviations

Abbreviation	Explanation
3R	„3R“ principles: Reduce, Replace, Refine of animal tests
AD	Applicability domain
BAF	Bioaccumulation factor
BCF	Bioconcentration factor
BMF	Biomagnification factor
CAS	Chemical Abstracts Service
DMSO	Dimethyl sulfoxide
ECHA	European Chemicals Agency
EU	European Union
FKZ	Forschungskennzahl (Project number)
HCH	Hexachlorocyclohexane
HEP	Hepatocytes
HYBIT	*Hyalella azteca* bioconcentration test
ITS	Integrated Testing Strategy
IME	Fraunhofer Institute for Molecular Biologie and Applied Ecology
IVIVE	In vitro - In vivo extrapolation
k_1	Uptake rate constant
k_2	Elimination rate constant
K_{OA}	Octanol-air partition coefficient
K_{OW}	Octanol-water partition coefficient
LSER	Linear solvation energy relationship
N/A	Not available
NAM	New Approach Methodologies
NPOC	Non-purgeable organic carbon
OECD	Organization for Economic Co-operation and Development
PAH	Polycyclic aromatic hydrocarbon
PBT	Persistent, bioaccumulating, toxic
PBTK	Physiologically Based Toxicokinetic
PCB	Polychlorinated biphenyl
PFAS	Per- and polyfluoroalkyl substances
PFOS	Perfluorooctane sulphonic acid

Abbreviation	Explanation
p*K*a	Acid dissociation constant
POP	Persistent Organic Pollutant
QSAR	Quantitative Structure Activity Relationship
REACH	Registration, Evaluation, Authorisation and Restriction of Chemicals
S9	Mixture of particularly prepared hepatic enzymes
TG	Test Guideline
TMF	Trophic magnification factor
TRR	Total radioactive residue(s)
TWA	Time-weighted average
UFZ	Helmholtz Centre for Environmental Research
UBA	Umweltbundesamt (German Environment Agency)
UVCB	Unknown or Variable composition, Complex reaction products or Biological materials
vPvB	Very persistent (vP) and very bioaccumulative (vB)
WAF	Water accommodated fraction
WoE	Weight of Evidence

Summary

The evaluation of bioaccumulation potential is one of the fundamental elements of environmental risk assessment for chemical regulation. Since it became evident that certain substances can accumulate in organisms to levels causing harm, bioaccumulation has remained a cornerstone of environmental protection frameworks. The historical roots of this awareness trace back to the environmental movement initiated by Rachel Carson's *Silent Spring* (1962), which prompted a global understanding of the adverse effects of persistent chemicals and inspired the development of legislative systems addressing environmental contamination.

Initially, bioaccumulation assessment concentrated on aquatic organisms, in particular fish, resulting in the Organisation for Economic Co-operation and Development (OECD) Test Guideline (TG) 305 (OECD 2012). Originally published as a set of several parts and later harmonised into a single guideline, TG 305 became the global standard for assessing the bioconcentration factor (BCF) in fish. Its revisions in 1996 and 2012 improved consistency in assessing chemical accumulation under controlled laboratory conditions. For many decades, the fish bioconcentration test provided the primary tool for identifying substances of concern. However, chemical regulation and scientific knowledge have advanced considerably. The criterion of bioaccumulation has been incorporated into international frameworks such as the Stockholm Convention on Persistent Organic Pollutants (POPs) and within the European Union's chemicals regulations. It is used for risk assessment and secondary poisoning assessment and is one of the criteria for Persistent, Bioaccumulative and Toxic (PBT) or very Persistent and very Bioaccumulative (vPvB) substances. The emergence of new classes of compounds, such as per- and polyfluoroalkyl substances (PFAS) and cyclic siloxanes, has introduced new challenges because their behaviour often deviates from conventional neutral, hydrophobic chemicals. Research has also demonstrated that bioaccumulation patterns in air-breathing organisms differ from those in aquatic species, revealing that certain substances may not be identified as bioaccumulative under traditional aquatic testing yet still accumulate in terrestrial mammals.

Simultaneously, societal and regulatory demands have shifted towards reducing or replacing vertebrate testing in line with the 3R principles (replacement, reduction and refinement of animal use) (Russel and Birch 1959). This principle has been endorsed through European legislation and by bodies such as the OECD. Consequently, alternative testing approaches, including *in vitro* systems, *in silico* modelling, and *invertebrate*-based assays, have been developed. New OECD guidelines such as TG 319 (in vitro substrate depletion assay) and TG 321 (*Hyalella azteca* bioconcentration test) now provide formalised methods for non-vertebrate and in vitro testing (OECD 2018a-c; OECD 2024a). The European Chemicals Agency (ECHA) has acknowledged these methods in its updated REACH guidance, yet their regulatory application remains limited due to restricted experience and uncertainty about their predictive reliability.

Different regulatory frameworks—REACH, Biocidal Products Regulation (BPR), Plant Protection Products Regulation (PPPR), and regulations concerning human and veterinary medicinal products—each require the assessment of bioaccumulation potential, yet they differ in criteria and data interpretation. Most frameworks continue to rely on fish-based BCFs, even though this measure may not adequately reflect the behaviour of polar, ionisable or complex substances. To address these discrepancies and to support modernisation of bioaccumulation assessment across European regulations, the German Environment Agency (UBA) initiated a comprehensive research project. The objectives were to evaluate the applicability, robustness and regulatory relevance of both established and alternative test systems, and to propose an integrated, tiered assessment strategy that minimises vertebrate testing without compromising environmental protection. Twenty-six substances covering a wide range of physicochemical properties were examined in OECD TG 305, TG 319 and TG 321 tests, and in some cases additional rat hepatocyte

assays. These tests provided comparative datasets enabling cross-method evaluation and exploration of each method's applicability domain.

Beyond laboratory testing, the project incorporated an extensive literature review and database analysis encompassing over 230 substances representing pharmaceuticals, biocides, UV filters, plasticisers, pesticides, surfactants and persistent pollutants. The analysis also included development of consensus models for estimating key parameters such as log K_{OW}, log K_{OA} and p*K*a, together with the evaluation of QSAR approaches for predicting BCF values. These datasets provided the empirical and theoretical foundation for developing an integrated bioaccumulation assessment strategy applicable to both aquatic and terrestrial organisms.

The overarching ambition is to shift from single-method animal testing to a multi-tiered, evidence-based evaluation combining in silico, in vitro, and invertebrate data, which allows the registrant to choose the method best suited to the respective chemical. By improving the understanding of non-animal methods, the project aims at reducing the need for animal testing while maintaining the high level of protection for humans and the environment. The results were discussed at a stakeholder workshop held in Berlin in 2025, involving representatives from regulatory authorities, industry, and academia, whose feedback informed the final structure and recommendations of this report.

Regulatory bioaccumulation assessment

Across frameworks, the regulatory landscape reveals consistent reliance on fish testing despite the availability of validated non-vertebrate and in vitro approaches. This reveals lack of certainty among the regulatory community and industry about applicability domains, advantages and disadvantages of the available New Approach Methodologies (NAMs), as well as vast differences among the EU chemical regulations in data requirements and the possibility to use other/additional data for hazard or risk assessment in a Weight of Evidence approach (WoE).

Review of existing integrated testing strategies for bioaccumulation assessment

Over the past two decades, several integrated or tiered approaches for bioaccumulation assessment have been developed to align with the 3R principles and improve regulatory efficiency. A seminal model by de Wolf et al. (2007) introduced a four-tier Integrated Testing Strategy (ITS) beginning with physicochemical analysis and computational prediction of ADME properties, followed by in vitro assays, modified in vivo studies, and culminating in a full OECD 305 test. The approach demonstrated that many chemicals could be adequately assessed without resorting to vertebrate testing, laying the groundwork for regulatory acceptance of alternative methods.

Subsequent proposals, such as that by Lombardo et al. (2014), refined the tiered system into a decision tree that permits waiving of in vivo tests for the majority of substances once sufficient in silico or in vitro evidence has been generated. According to this study, approximately two-thirds of the assessed substances could be concluded without vertebrate testing, illustrating substantial potential for reduction.

Parallel developments addressed bioaccumulation in air-breathing organisms. In 2022, the ECHA published a discussion paper proposing a formal assessment scheme for mammals, emphasising the use of multiple lines of evidence including in silico predictions and in vitro data on metabolism. This marked a crucial step towards acknowledging bioaccumulation pathways beyond lipid partitioning in aquatic systems (ECHA, 2022)

The Bioaccumulation Assessment Tool (BAT) (Armitage et al. 2021; Arnot et al. 2023) provides a framework for integrating and weighting diverse datasets through a quantitative WoE approach.

Implemented in a spreadsheet interface, BAT allows systematic evaluation of data reliability and coherence across in vivo, in vitro, and in silico sources. Although powerful, its application in regulatory decision-making remains limited. While the BAT-Tool standardises evidence evaluation, a key limitation is that in regulatory practice "reliable evidence of bioaccumulation cannot be outweighed by information showing no bioaccumulation" (ECHA 2023a).

Most recently, the OECD has introduced an Integrated Approach to Testing and Assessment (IATA) for bioaccumulation (OECD 2024c). This framework formalises the tiered, iterative process through which data from different sources are combined, uncertainties evaluated, and evidence weighted according to relevance and reliability. The IATA structure emphasises progression from lower-tier, resource-efficient methods to higher-tier studies only where necessary, promoting both ethical and scientific consistency. In contrast with traditional regulatory schemes that depend on single definitive endpoints such as BCF from fish testing, the OECD IATA provides a transparent mechanism for integrating varied data types. It explicitly addresses uncertainty, encourages flexibility, and supports cross-regulatory harmonisation. Nevertheless, differences in jurisdictional thresholds and interpretative standards still hinder universal adoption.

A central conclusion from reviewing existing ITS and IATA developments is that bioaccumulation assessment must move towards scientifically coherent, animal-sparing strategies applicable across all sectors. Future frameworks must accommodate proposals how to deal with substances with challenging properties, including ionisable, highly hydrophobic or amphiphilic compounds, and how to account for non-lipid accumulation mechanisms such as sorption processes.

Proposal for an integrated assessment strategy for aquatic and terrestrial bioaccumulation

Building upon existing testing strategies and regulatory guidance, the project developed a comprehensive integrated assessment strategy for evaluating bioaccumulation in both aquatic and terrestrial organisms. The framework encompasses multiple tiers that combine in silico, in vitro, and invertebrate testing with established in vivo data, aiming to minimise animal use while maintaining regulatory reliability. It covers classical neutral hydrophobic substances as well as ionisable, superhydrophobic and amphiphilic compounds. For new substances without prior data, the scheme guides stepwise testing to obtain the information necessary for decision-making. For existing substances, all available and relevant information, including monitoring data and previous studies, is integrated into a WoE evaluation before determining whether additional tests would be required for reliable decision-making.

The framework encompasses two assessment schemes, an aquatic scheme and a parallel tiered system for air-breathing organisms. Tier 1 of the aquatic scheme uses hydrophobicity as the primary physicochemical indicator. The octanol–water partition coefficient (log K_{OW}) is used as a surrogate for bioaccumulation potential, with consensus modelling recommended to improve reliability. For ionisable substances, the distribution coefficient (log D) at relevant environmental pH values is used instead. Regulatory thresholds vary by framework, but generally a log $K_{OW} < 2$ indicates negligible bioaccumulation. Values above 3 trigger further evaluation of the aquatic bioaccumulation potential. A log K_{OW} between 3 and 6 triggers prediction of the BCF using QSAR models in tier 2. For superhydrophobic substances (tentatively defined as log $K_{OW} > 6$) experimental testing may become unreliable due to sorption artefacts, thus a precautionary assumption of bioaccumulation is recommended unless evidence of rapid biotransformation is available.

Tier 2 employs QSAR models to estimate BCF values. Consolidation of at least four independent predictions is advised to account for model variability. When modelled BCFs are well below

thresholds and within model applicability domains, testing may be waived. Otherwise, results proceed to tier 3 involving in vitro or non-vertebrate assays.

Tier 3 integrates the in vitro substrate depletion assays (OECD 319 A/B), which estimate biotransformation rates in fish liver S9 fractions or hepatocytes and the *Hyalella azteca* bioconcentration test (OECD 321). Either one of these methods offer a mechanistic understanding of uptake and metabolism without the need for live fish testing. When combined with in silico predictions, they can often provide sufficient confidence for regulatory decision-making. Only when substantial uncertainty remains, or where thresholds are narrowly exceeded, should tier 4 be invoked, with fish studies according to OECD TG 305. All other data, such as field and monitoring data to confirm long-term accumulation patterns should also be considered in a WoE approach, to support regulatory decisions. This was not separately evaluated in this project.

The integrated scheme extends beyond lipid-based accumulation to address substances accumulating via alternative mechanisms. For ionisable compounds, PFAS, or surfactants, supplementary assessments based on specific binding, sorption, or amphiphilic partitioning are incorporated. The strategy also links aquatic bioaccumulation assessment with considerations for secondary poisoning, ensuring consistent interpretation of exposure across trophic levels.

For air-breathing organisms, a parallel tiered system is proposed for cases where a substance does not accumulate in aquatic organisms but is expected to bioaccumulate in air-breathing organisms. Here, integrating partition coefficients such as log K_{OA} and metabolic capacity data to predict accumulation in terrestrial mammals and birds are proposed. The overall aim is to provide a coherent, cross-compartmental framework that aligns with regulatory criteria while reducing reliance on vertebrate testing.

Evaluation of test systems

A detailed evaluation of established and alternative methods for determining the bioaccumulation potential of chemicals was carried out as part of the project. The assessment compared traditional fish-based in vivo bioconcentration testing (OECD TG 305) with alternative NAMs, including the *Hyalella azteca* bioconcentration test (OECD TG 321), in vitro substrate depletion assays using trout hepatocytes or S9 fractions (OECD TG 319A/B), in vitro biotransformation tests with rat hepatocytes or liver S9 fractions, quantitative structure–activity relationship (QSAR) modelling, and the determination of physicochemical parameters such as log K_{OW}, log K_{OA}, and p*K*a. The objective was to assess the consistency, applicability domains, and regulatory relevance of these methods and to identify their limitations.

Data collection was based on comprehensive literature research and targeted laboratory testing. The dataset included 231 chemicals across diverse substance classes, such as persistent organic pollutants (POPs), polychlorinated biphenyls (PCBs), polycyclic aromatic hydrocarbons (PAHs), flame retardants, pesticides, pharmaceuticals, fragrances, biocides, UV filters, plasticisers, siloxanes, per- and polyfluoroalkyl substances (PFAS), and surfactants. For 26 of these substances, nearly complete sets of data were obtained across all methods, forming the basis for comparative evaluation.

The *Hyalella azteca* bioconcentration test (OECD TG 321) provided valuable information on the bioconcentration potential of substances in a non-vertebrate species. The test proved technically feasible for both neutral and ionisable substances, though further refinement of exposure conditions was recommended for highly hydrophobic or pH-sensitive chemicals. Results confirmed that the amphipod assay offers a suitable alternative for aquatic bioaccumulation assessment under regulatory frameworks, aligning with recent decisions to accept the method under REACH.

The in vitro substrate depletion assays according to OECD TG 319A/B yielded consistent estimates of biotransformation rates across a log K_{OW} range of 2–6. These rates were converted to predicted in vivo BCF values through in vitro–in vivo extrapolation (IVIVE) using established models such as B-compass fish. Trout hepatocyte and S9 assays demonstrated reliable identification of metabolically active versus stable substances. For example, pyrene was rapidly metabolised, whereas β-HCH and UV-234 showed limited transformation. The assays were fast, resource-efficient, and required small substance quantities, but technical challenges were encountered for volatile and highly hydrophobic substances. Their successful application highlighted their potential for screening purposes and integration into regulatory testing strategies.

In vitro rat liver hepatocyte assays were introduced to assess biotransformation in air-breathing organisms. Depletion rates obtained from these assays enable the estimation of biomagnification factors (BMFs) using IVIVE models, providing mechanistic insights into mammalian metabolism. Pyrene and chlorpyrifos were readily metabolised, whereas UV-234 was not. While results demonstrated the general suitability of hepatocytes for deriving clearance rates, further work was recommended to harmonise protocols and to validate their application for mammalian species.

QSAR modelling of BCFs was conducted using a range of models available, for example, in EpiSuite, VEGA, and the OCHEM. Variability among model outputs was notable, with differences of more than one log unit observed for some substances. Consensus modelling, integrating multiple model predictions and measured data, was identified as the most robust approach. For hydrophilic compounds (log K_{OW} < 3), QSAR-predicted BCFs reliably indicated low bioaccumulation, while results for superhydrophobic compounds were inconsistent. Consolidation of QSAR results across datasets allowed estimation of uncertainty ranges and improved predictive reliability. However, for regulatory use, further validation of these QSARs is necessary.

The evaluation of physicochemical parameters revealed similar challenges. The determination of log K_{OW} using different experimental and computational methods produced substantial variability. Iterative consensus modelling was found to reduce uncertainty and to provide more stable input values for bioaccumulation assessment. For log K_{OA}, data scarcity remained a limiting factor, though a similar consensus approach was recommended. The estimation of acid dissociation constants (p*K*a) was crucial for assessing ionisable substances. Comparative evaluation of estimation tools such as ACD/Labs GALAS and OPERA 2.6 demonstrated acceptable agreement for acids but greater scatter for bases. The study concluded that consistent p*K*a values obtained from at least two independent methods fulfil OECD criteria for regulatory use.

Overall, it was confirmed that NAMs, QSARs, and refined physicochemical parameters can jointly provide reliable information for excluding bioaccumulation concern under defined conditions. However, their application is limited for superhydrophobic, ionisable and substances not accumulating predominantly in lipid, which may require case-specific adaptations or complementary testing.

Evaluation of the integrated test and assessment strategy

The evaluation of the integrated assessment strategy focused on examining the coherence, reliability, and predictive performance of the combined methods for 26 case study substances. The analysis compared datasets from the *Hyalella azteca* test (OECD TG 321), in vitro assays (OECD TG 319A/B), in vivo fish BCF studies (OECD TG 305), and QSAR predictions. The aim was to determine whether the integrated, tiered approach provides a scientifically sound basis for regulatory decision-making while reducing the need for vertebrate testing.

For substances with low hydrophobicity, physicochemical screenings and QSAR estimates were sufficient to determine non-bioaccumulation, meaning that theoretically no further testing was necessary. However, for moderately to highly hydrophobic substances, discrepancies between HYBIT, in vitro and in vivo data were observed for certain chemicals, underscoring the importance of expert interpretation of the results. For highly hydrophobic substances, the use of currently available NAMs was limited due to methodological difficulties. The study highlights the need for targeted strategies and further research for these substances.

The evaluation also addressed substances not predominantly accumulating in lipid phases, including ionisable compounds, PFAS, and surfactants. Ionisable substances were assessed using log D and p*K*a-dependent distribution models, while PFAS and surfactants required specific consideration of binding and partitioning behaviour. Current test methods provided preliminary insights but were not universally applicable, underlining the need for further methodological development.

For air-breathing organisms, in vitro biotransformation data from rat hepatocyte assays showed potential, but the estimation of bioaccumulation potential through IVIVE models needs to be further developed. While still in an early stage of validation, the approach using rat in vitro assays demonstrated strong potential for assessing biomagnification without animal testing. The proposed tier 2 threshold—where a depletion rate constant exceeding 3 h^{-1} in rat S9 assays indicates low bioaccumulation—was found to be a practical screening criterion, though further refinement and adaptation to rat hepatocytes is needed.

In summary, the integrated testing and assessment strategy proved effective for identifying substances of low bioaccumulation concern and for guiding efficient testing sequences. It confirmed the complementary roles of in silico, in vitro, and invertebrate methods and highlighted the importance of expert judgement for complex cases. NAMs can substantially reduce vertebrate testing while maintaining regulatory protection levels.

Concluding reflections and future pathways

The integrated assessment scheme presented in this project provides a comprehensive and forward-looking framework for bioaccumulation assessment across chemical regulations. It enables regulators to move beyond outdated reliance on vertebrate testing, supports the assessment of the broad variety of chemical substances, and promotes cross-regulatory coherence. Built on a flexible five-tier structure, the integrated assessment scheme emphasises early screening, use of in silico data and NAMs, to ensure minimal animal testing. The study shows that NAMs can effectively exclude bioaccumulation concerns under certain conditions, reducing reliance on animal tests. Yet, for challenging substances, critical review and complementary methods remain essential. In the context of growing environmental pressures, technological capability, and societal expectations, the integrated assessment scheme offers a concept for the next generation of chemical safety assessment—one that is evidence-based, efficient, and ethically sound. Its adoption will not only enhance the scientific quality of regulatory decisions but also contribute to a more sustainable and protective approach to environmental health.

Zusammenfassung

Die Bewertung des Bioakkumulationspotenzials ist eines der grundlegenden Elemente der Umweltrisikobewertung für die Chemikalienregulierung. Seitdem bekannt ist, dass sich bestimmte Stoffe in Organismen in schädlichen Konzentrationen anreichern können, ist die Bioakkumulation ein Eckpfeiler des Umweltschutzes geblieben. Die historischen Wurzeln dieses Bewusstseins reichen zurück bis zur Umweltbewegung, die durch Rachel Carsons Buch „Der stumme Frühling" (1962) ausgelöst wurde. Dieses Buch führte zu einem globalen Verständnis der schädlichen Auswirkungen persistenter Chemikalien und inspirierte die Entwicklung von Rechtssystemen zur Bekämpfung der Umweltverschmutzung.

Anfangs konzentrierte sich die Bewertung der Bioakkumulation auf Wasserorganismen, insbesondere Fische, was zur Testrichtlinie (TG) 305 der Organisation für wirtschaftliche Zusammenarbeit und Entwicklung (OECD) führte (OECD 2012). Ursprünglich als mehrteilige Reihe veröffentlicht und später zu einer einzigen Leitlinie harmonisiert, wurde TG 305 zum globalen Standard für die Bewertung des Biokonzentrationsfaktors (BCF) in Fischen. Durch Überarbeitungen der Testrichtlinie in den Jahren 1996 und 2012 wurde die Konsistenz bei der Bewertung der Anreicherung von Chemikalien unter kontrollierten Laborbedingungen verbessert. Viele Jahrzehnte lang war der Fisch-Biokonzentrationstest das wichtigste Instrument zur Identifizierung bedenklicher Stoffe. Seitdem haben sich jedoch die Chemikalienregulierung und die wissenschaftlichen Erkenntnisse erheblich weiterentwickelt. Das Kriterium der Bioakkumulation wurde in internationale Rahmenwerke wie das Stockholmer Übereinkommen über persistente organische Schadstoffe (POP) und in die Chemikalienvorschriften der Europäischen Union aufgenommen. Es wird für die Risikobewertung und die Bewertung von Sekundärvergiftungen verwendet und ist eines der Kriterien für persistente, bioakkumulierbare und toxische (PBT) oder sehr persistente und sehr bioakkumulierbare (vPvB) Stoffe. Das Aufkommen neuer Klassen von Verbindungen, wie per- und polyfluorierte Alkylsubstanzen (PFAS) und cyclische Siloxane, hat neue Herausforderungen mit sich gebracht, da ihr Verhalten oft von dem herkömmlicher neutraler, lipophiler Chemikalien abweicht. Untersuchungen haben auch gezeigt, dass sich die Bioakkumulationsmuster bei luftatmenden Organismen von denen bei aquatischen Arten unterscheiden, was darauf hindeutet, dass bestimmte Stoffe unter traditionellen aquatischen Testbedingungen möglicherweise nicht als bioakkumulierbar identifiziert werden, sich aber dennoch in Landwirbeltieren anreichern.

Gleichzeitig haben sich die gesellschaftlichen und regulatorischen Anforderungen dahingehend verändert, dass Versuche an Wirbeltieren gemäß den 3R-Prinzipien (Ersatz, Verringerung und Verfeinerung der Tiernutzung) reduziert oder ersetzt werden sollen (Russel and Burch 1959). Dieses Prinzip wurde durch europäische Rechtsvorschriften und von Gremien wie der OECD bestätigt. Infolgedessen wurden alternative Testansätze entwickelt, darunter In-vitro-Systeme, In-silico-Modellierung und Tests mit Wirbellosen. Neue OECD-Leitlinien wie TG 319A/B (In-vitro-Biotransformation) und TG 321 (*Hyalella azteca*-Biokonzentrationstest) bieten nun formalisierte Methoden für Tests an Wirbellosen und In-vitro-Tests (OECD 2018a-c; OECD 2024a). Die Europäische Chemikalienagentur (ECHA) hat diese Methoden in ihren aktualisierten REACH-Leitlinien anerkannt, doch ihre regulatorische Anwendung bleibt aufgrund begrenzter Erfahrungen und Unsicherheiten hinsichtlich ihrer Vorhersagbarkeit weiterhin eingeschränkt.

Verschiedene Rechtsrahmen – REACH, Biozidprodukteverordnung (BPR), Pflanzenschutzmittelverordnung (PPPR) und Vorschriften für Human- und Tierarzneimittel – verlangen jeweils die Bewertung des Bioakkumulationspotenzials, unterscheiden sich jedoch hinsichtlich der Kriterien und der Datenauswertung. Die meisten Rechtsrahmen stützen sich weiterhin auf fischbasierte BCFs, obwohl diese Messgröße das Verhalten polarer, ionisierbarer oder komplexer Stoffe möglicherweise nicht angemessen widerspiegelt.

Um diese Diskrepanzen zu beseitigen und die Modernisierung der Bioakkumulationsbewertung in den europäischen Vorschriften zu unterstützen, hat das Umweltbundesamt (UBA) ein umfassendes Forschungsprojekt initiiert. Ziel war es, die Anwendbarkeit, Robustheit und regulatorische Relevanz sowohl etablierter als auch alternativer Testsysteme zu bewerten und eine integrierte, mehrstufige Bewertungsstrategie vorzuschlagen, die Versuche an Wirbeltieren minimiert, ohne den Umweltschutz zu beeinträchtigen. Sechsundzwanzig Substanzen mit einem breiten Spektrum an physikalisch-chemischen Eigenschaften wurden in den OECD-Tests TG 305, TG 319 und TG 321 und in einigen Fällen zusätzlich in Rattenhepatozyten-Assays untersucht. Diese Tests lieferten vergleichbare Datensätze, die eine methodenübergreifende Bewertung und Untersuchung des Anwendungsbereichs jeder Methode ermöglichten.

Über die Labortests hinaus umfasste das Projekt eine umfassende Literaturrecherche und Datenbankanalyse, die über 230 Substanzen aus den Bereichen Arzneimittel, Biozide, UV-Filter, Weichmacher, Pestizide, Tenside und persistente Schadstoffe umfasste. Die Analyse umfasste auch die Entwicklung von Konsensmodellen zur Abschätzung wichtiger Parameter wie log K_{OW}, log K_{OA} und p*K*a sowie die Bewertung von QSAR-Ansätzen zur Vorhersage von BCF-Werten. Diese Datensätze bildeten die empirische und theoretische Grundlage für die Entwicklung einer integrierten Strategie zur Bewertung der Bioakkumulation, die sowohl für aquatische als auch für terrestrische Organismen anwendbar ist.

Das übergeordnete Ziel besteht darin, von Tierversuchen mit einer einzigen Methode zu einer mehrstufigen, evidenzbasierten Bewertung überzugehen, bei der In-silico-, In-vitro- und Wirbellosen-Daten kombiniert werden, sodass der Registrant die für die jeweilige Chemikalie am besten geeignete Methode auswählen kann. Durch ein besseres Verständnis tierversuchsfreier Methoden soll das Projekt eine weitere Harmonisierung der europäischen Rechtsvorschriften ermöglichen und gleichzeitig ein hohes Schutzniveau für Mensch und Umwelt aufrechterhalten sowie den ethischen und wissenschaftlichen Anforderungen an die Nachhaltigkeit gerecht werden. Die Ergebnisse wurden 2025 auf einem Stakeholder-Workshop in Berlin diskutiert, an dem Vertreter von Regulierungsbehörden, Industrie und Wissenschaft teilnahmen, deren Feedback in die endgültige Struktur und die Empfehlungen dieses Berichts einfloss.

Regulatorische Bewertung der Bioakkumulation

Über alle Rahmenwerke hinweg zeigt sich, dass die Regulierungsbehörden trotz der Verfügbarkeit validierter Nicht-Wirbeltier- und In-vitro-Ansätze weiterhin konsequent auf Fischversuche setzen. Dies verdeutlicht die Unsicherheit der Regulierungsbehörden und der Industrie hinsichtlich der Anwendungsbereiche, der Vor- und Nachteile der verfügbaren NAMs sowie der großen Unterschiede zwischen den EU-Chemikalienvorschriften in Bezug auf die Datenanforderungen und die Möglichkeit, andere/zusätzliche Daten für die Gefahren- oder Risikobewertung im Rahmen eines Weight-of-Evidence-Ansatzes (WoE) zu verwenden.

Überprüfung bestehender integrierter Teststrategien für die Bewertung der Bioakkumulation

In den letzten zwei Jahrzehnten wurden mehrere integrierte oder mehrstufige Ansätze für die Bewertung der Bioakkumulation entwickelt, um sie an die 3R-Prinzipien anzupassen und die Effizienz der Regulierung zu verbessern. Ein wegweisendes Modell von de Wolf et al. (2007) führte eine vierstufige integrierte Teststrategie (ITS) ein, die mit einer physikalisch-chemischen Analyse und einer computergestützten Vorhersage der ADME-Eigenschaften beginnt, gefolgt von In-vitro-Assays, modifizierten In-vivo-Studien und schließlich einem vollständigen OECD-305-Test. Der Ansatz zeigte, dass viele Chemikalien ohne Rückgriff auf Versuche an Wirbeltieren angemessen bewertet werden können, und legte damit den Grundstein für die regulatorische Akzeptanz alternativer Methoden.

Spätere Vorschläge, wie beispielsweise der von Lombardo et al. (2014), verfeinerten das mehrstufige System zu einem Entscheidungsbaum, der es ermöglicht, In-vivo-Tests für die meisten Stoffe zu vermeiden, sobald ausreichende In-silico- oder In-vitro-Nachweise vorliegen. Bei etwa zwei Dritteln der bewerteten Stoffe konnte auf Versuche an Wirbeltieren verzichtet werden, was ein erhebliches Reduktionspotenzial verdeutlicht.

Parallel dazu wurden Entwicklungen im Bereich der Bioakkumulation in luftatmenden Organismen vorangetrieben. Im Jahr 2022 veröffentlichte die Europäische Chemikalienagentur ein Diskussionspapier, in dem ein formelles Bewertungsschema für Säugetiere vorgeschlagen wurde, wobei der Schwerpunkt auf der Verwendung mehrerer Beweislinien lag, darunter In-silico-Vorhersagen und In-vitro-Daten zum Stoffwechsel. Dies war ein entscheidender Schritt zur Anerkennung von Bioakkumulationswegen, die über die Lipidverteilung in aquatischen Systemen hinausgehen (ECHA 2022).

Das Bioaccumulation Assessment Tool (BAT) bietet einen Rahmen für die Integration und Gewichtung verschiedener Datensätze durch einen quantitativen WoE-Ansatz (Armitage et al. 2021; Arnot et al. 2023). Das BAT wird in einer Dateneingabeoberfläche implementiert und ermöglicht eine systematische Bewertung der Datenzuverlässigkeit und -kohärenz aus In-vivo-, In-vitro- und In-silico-Quellen. Obwohl es sehr leistungsfähig ist, bleibt seine Anwendung in der regulatorischen Entscheidungsfindung begrenzt. Das BAT-Tool standardisiert zwar die Bewertung von Nachweisen, eine wesentliche Einschränkung besteht jedoch darin, dass in der Regulierungspraxis „zuverlässige Nachweise für eine Bioakkumulation nicht durch Informationen, die keine Bioakkumulation belegen, aufgewogen werden können“ (ECHA 2023a).

Vor kurzem hat die OECD einen integrierten Ansatz für die Prüfung und Bewertung (IATA) der Bioakkumulation veröffentlicht (OECD 2024c). Dieser Rahmen formalisiert den mehrstufigen, iterativen Prozess, durch den Daten aus verschiedenen Quellen kombiniert, Unsicherheiten bewertet und Beweise nach Relevanz und Zuverlässigkeit gewichtet werden. Die IATA-Struktur betont den Übergang von ressourceneffizienten Methoden der unteren Stufe zu Studien der höheren Stufe nur dann, wenn dies erforderlich ist, und fördert damit sowohl ethische als auch wissenschaftliche Konsistenz. Im Gegensatz zu traditionellen Regulierungssystemen, die sich auf einzelne definitive Endpunkte wie die BCF aus Fischversuchen stützen, bietet der IATA der OECD einen transparenten Mechanismus für die Integration verschiedener Datentypen. Er geht ausdrücklich auf Unsicherheiten ein, fördert Flexibilität und unterstützt die Harmonisierung über verschiedene Rechtsvorschriften hinweg. Dennoch behindern Unterschiede in den rechtlichen Schwellenwerten und Auslegungsstandards nach wie vor eine universelle Anwendung.

Eine zentrale Schlussfolgerung aus der Überprüfung bestehender ITS- und IATA-Entwicklungen ist, dass die Bewertung der Bioakkumulation in Richtung wissenschaftlich kohärenter, tierversuchsfreier Strategien gehen muss, die in allen Sektoren anwendbar sind. Zukünftige Rahmenwerke müssen Vorschläge enthalten, wie mit Substanzen mit schwierigen Eigenschaften, einschließlich ionisierbarer, stark hydrophober oder amphiphiler Verbindungen, umgegangen werden soll und wie Nicht-Lipid-Akkumulationsmechanismen wie Sorptionsprozesse berücksichtigt werden können.

Vorschlag für eine integrierte Bewertungsstrategie für die Bioakkumulation in Gewässern und auf dem Land

Aufbauend auf bestehenden Teststrategien und regulatorischen Leitlinien wurde im Rahmen des Projekts eine umfassende integrierte Bewertungsstrategie zur Bewertung der Bioakkumulation in Wasser- und Landorganismen entwickelt. Der Rahmen umfasst mehrere Ebenen, die In-silico-, In-vitro- und Wirbellosen-Tests mit etablierten In-vivo-Daten kombinieren, um den Einsatz von Tieren zu minimieren und gleichzeitig die regulatorische

Zuverlässigkeit zu gewährleisten. Das Schema umfasst klassische neutrale lipophile Substanzen sowie ionisierbare, superhydrophobe und amphiphile Verbindungen. Für neue Substanzen, für die keine Daten vorliegen, sieht das Schema schrittweise Tests vor, um die für die Entscheidungsfindung erforderlichen Informationen zu erhalten. Bei bestehenden Substanzen werden alle verfügbaren und relevanten Informationen, einschließlich Monitoringdaten und die Ergebnisse früherer Studien, in eine WoE-Bewertung einbezogen, bevor entschieden wird, ob zusätzliche Tests für eine zuverlässige Entscheidungsfindung erforderlich sind.

Das Bewertungssystem umfasst zwei Schemata, ein Schema für aquatische Organismen und ein paralleles System für luftatmende Wirbeltiere. In Stufe 1 des aquatischen Schemas wird die Hydrophobie als primärer physikalisch-chemischer Indikator verwendet. Der Oktanol-Wasser-Verteilungskoeffizient (log K_{OW}) wird als Maß für das Bioakkumulationspotenzial verwendet, wobei zur Verbesserung der Zuverlässigkeit eine Konsensmodellierung empfohlen wird. Für ionisierbare Stoffe wird stattdessen der Verteilungskoeffizient (log D) bei relevanten pH-Werten in der Umwelt verwendet. Die regulatorischen Schwellenwerte variieren je nach Rahmenwerk, aber im Allgemeinen weist ein log $K_{OW} < 2$ auf eine vernachlässigbare Bioakkumulation hin, während Werte über 3 eine weitere Bewertung auslösen. Bei einem log K_{OW} zwischen 3 und 6 stellt die Vorhersage des BCF unter Verwendung von QSAR-Modellen die Stufe 2 dar. Bei stark hydrophoben Stoffen (vorläufig definiert als log $K_{OW} > 6$) können experimentelle Tests aufgrund von Sorptionsartefakten unzuverlässig werden, sodass eine vorsorgliche Annahme der Bioakkumulation empfohlen wird, sofern keine Hinweise auf eine schnelle Biotransformation vorliegen.

In Stufe 2 werden QSAR-Modelle zur Schätzung der BCF-Werte verwendet. Es wird empfohlen, mindestens vier unabhängige Vorhersagen zu konsolidieren, um die Modellvariabilität zu berücksichtigen. Wenn die modellierten BCF-Werte deutlich unter den Schwellenwerten und innerhalb der Anwendbarkeitsbereiche des Modells liegen, kann auf Tests verzichtet werden. Andernfalls werden die Ergebnisse in Stufe 3 weiterverarbeitet, die Tests an Wirbellosen oder In-vitro-Tests umfasst.

Stufe 3 umfasst den *Hyalella azteca*-Biokonzentrationstest (OECD TG 321) und die In-vitro-Substratabbauversuche (OECD TG 319A/B), mit denen die Biotransformationsraten in S9-Fraktionen oder Hepatozyten der Fischleber geschätzt werden. Beide Methoden ermöglichen ein mechanistisches Verständnis der Aufnahme und des Stoffwechsels, ohne dass Versuche mit lebenden Fischen erforderlich sind. In Kombination mit In-silico-Vorhersagen können sie oft ausreichende Sicherheit für regulatorische Entscheidungen bieten. Nur wenn erhebliche Unsicherheiten bestehen bleiben oder Grenzwerte knapp unterschritten sind, sollte Tier 4 mit Fischstudien gemäß OECD TG 305 herangezogen werden. Alle anderen Daten, wie Feld- und Monitoringdaten zur Bestätigung langfristiger Akkumulationsmuster, sollten ebenfalls in einem WoE-Ansatz berücksichtigt werden, um regulatorische Entscheidungen zu unterstützen. Dies wurde in diesem Projekt nicht separat bewertet.

Das integrierte Konzept geht über die Anreicherung auf Lipidbasis hinaus und befasst sich auch mit Substanzen, die sich über alternative Mechanismen anreichern. Für ionisierbare Verbindungen, PFAS oder Tenside werden ergänzende Bewertungen auf der Grundlage von Sorption, amphiphiler Verteilung oder spezifischer Bindung einbezogen. Das integrierte Konzept verbindet auch die Bewertung der Bioakkumulation in Gewässern mit Überlegungen zur Sekundärvergiftung (Secondary poisoning) und gewährleistet so eine einheitliche Interpretation der Exposition über alle trophischen Ebenen hinweg.

Für luftatmende Organismen wird ein paralleles Stufensystem für Fälle vorgeschlagen, in denen sich eine Substanz nicht in aquatischen Organismen anreichert, aber eine Bioakkumulation in luftatmenden Wirbeltieren zu erwarten ist. Hier wird vorgeschlagen, Verteilungskoeffizienten

wie log K_{OA} und Daten zur Stoffwechselkapazität zu integrieren, um die Anreicherung in Landwirbeltieren und Vögeln vorherzusagen. Das übergeordnete Ziel besteht darin, einen kohärenten, kompartimentübergreifenden Rahmen zu schaffen, der mit den regulatorischen Kriterien im Einklang steht und gleichzeitig die Abhängigkeit von Versuchen an Wirbeltieren verringert.

Bewertung von Testsystemen

Im Rahmen des Projekts wurde eine detaillierte Bewertung etablierter und alternativer Methoden zur Bestimmung des Bioakkumulationspotenzials von Chemikalien durchgeführt. Bei der Bewertung wurden traditionelle In-vivo-Biokonzentrationstests mit Fischen (OECD TG 305) mit alternativen Methoden (NAMs) verglichen, darunter der *Hyalella azteca*-Biokonzentrationstest (OECD TG 321), In-vitro-Biotransformationstests mit Forellenhepatozyten oder S9-Fraktionen (OECD TG 319A/B), In-vitro-Biotransformationstests mit Rattenhepatozyten oder Leber-S9-Fraktionen, quantitative Struktur-Wirkungs-Beziehungsmodelle (QSAR) und die Bestimmung physikalisch-chemischer Parameter wie log K_{OW}, log K_{OA} und p*K*a. Ziel war es, die Konsistenz, Anwendungsbereiche und regulatorische Relevanz dieser Methoden zu bewerten und ihre Grenzen zu identifizieren.

Die Datenerhebung basierte auf einer umfassenden Literaturrecherche und gezielten Labortests. Der Datensatz umfasste 231 Chemikalien aus verschiedenen Stoffklassen, darunter persistente organische Schadstoffe (POP), polychlorierte Biphenyle (PCB), polyzyklische aromatische Kohlenwasserstoffe (PAK), Flammschutzmittel, Pestizide, Arzneimittel, Duftstoffe, Biozide, UV-Filter, Weichmacher, Siloxane, per- und polyfluorierte Alkylsubstanzen (PFAS) und Tenside. Für 26 dieser Stoffe wurden nahezu vollständige Datensätze für alle Methoden erhoben, die die Grundlage für eine vergleichende Bewertung bildeten.

Der *Hyalella azteca*-Biokonzentrationstest (OECD TG 321) lieferte wertvolle Informationen über das Biokonzentrationspotenzial von Stoffen in einer wirbellosen Tierart. Der Test erwies sich sowohl für neutrale als auch für ionisierbare Substanzen als technisch durchführbar, allerdings wurde eine weitere Verfeinerung der Expositionsbedingungen für stark hydrophobe oder pH-empfindliche Chemikalien empfohlen. Die Ergebnisse bestätigten, dass der Amphipoden-Test eine geeignete Alternative für die Bewertung der Bioakkumulation in aquatischen Organismen im Rahmen der gesetzlichen Vorschriften darstellt, was mit den jüngsten Entscheidungen zur Zulassung der Methode im Rahmen von REACH im Einklang steht.

Die In-vitro-Biotransformationstests gemäß OECD TG 319A/B ergaben konsistente Schätzungen der Biotransformationsraten über einen log K_{OW} -Bereich von 2 – 6. Diese Raten wurden durch In-vitro/In-vivo-Extrapolation (IVIVE) unter Verwendung etablierter Modelle wie B-Compass Fisch in In-vivo-BCF-Werte umgerechnet. Forellenhepatozyten- und S9-Assays ermöglichten eine zuverlässige Identifizierung von metabolisch aktiven gegenüber stabilen Substanzen. So wurde beispielsweise Pyren schnell metabolisiert, während β-HCH und UV-234 nur eine begrenzte Umwandlung zeigten. Die Assays waren schnell, ressourceneffizient und erforderten nur geringe Substanzmengen, jedoch traten bei flüchtigen und stark hydrophoben Substanzen technische Herausforderungen auf. Ihre erfolgreiche Anwendung unterstreicht ihr Potenzial für Screening-Zwecke und die Integration in regulatorische Teststrategien.

In-vitro-Assays mit Rattenleberhepatozyten wurden eingeführt, um die Biotransformation in luftatmenden Organismen zu bewerten. Anhand der aus diesen Tests gewonnenen Abbauraten lassen sich mithilfe von IVIVE-Modellen Biomagnifikationsfaktoren (BMFs) abschätzen, die mechanistische Einblicke in den Stoffwechsel von Säugetieren liefern. Pyren und Chlorpyrifos wurden leicht metabolisiert, UV-234 hingegen nicht. Die Ergebnisse zeigten zwar die allgemeine Eignung von Hepatozyten für die Ableitung von Clearance-Raten, jedoch wurden weitere

Arbeiten empfohlen, um die Protokolle zu harmonisieren und ihre Anwendung für Säugetierarten zu validieren.

Die QSAR-Modellierung der BCFs wurde unter Verwendung einer Reihe von Modellen durchgeführt, die zum Beispiel in EpiSuite, VEGA und OCHEM verfügbar sind. Die Variabilität zwischen den Modellergebnissen war bemerkenswert, wobei bei einigen Stoffen Unterschiede von mehr als einer Log-Einheit beobachtet wurden. Die Konsensmodellierung, bei der mehrere Modellvorhersagen und Messdaten integriert wurden, wurde als der robusteste Ansatz identifiziert. Bei hydrophilen Verbindungen (log K_{OW} < 3) deuteten die QSAR-basierte BCFs zuverlässig auf eine geringe Bioakkumulation hin, während die Ergebnisse für superhydrophobe Verbindungen inkonsistent waren. Die Konsolidierung der QSAR-Ergebnisse über verschiedene Datensätze hinweg ermöglichte die Abschätzung von Unsicherheitsbereichen und verbesserte die Vorhersagegenauigkeit. Für regulatorische Zwecke ist jedoch eine weitere Validierung dieser QSARs erforderlich.

Die Bewertung der physikalisch-chemischen Parameter ergab ähnliche Herausforderungen. Die Bestimmung von log K_{OW} unter Verwendung verschiedener experimenteller und rechnerischer Methoden führte zu erheblichen Schwankungen. Es wurde festgestellt, dass iterative Konsensmodellierung die Unsicherheit verringert und stabilere Eingabewerte für die Bewertung der Bioakkumulation liefert. Für log K_{OA} blieb die Datenknappheit ein limitierender Faktor, obwohl ein ähnlicher Konsensansatz empfohlen wurde. Die Schätzung der Säurekonstanten (p*K*a) war für die Bewertung ionisierbarer Stoffe von entscheidender Bedeutung. Eine vergleichende Bewertung von Schätzwerkzeugen wie ACD/Labs GALAS und OPERA 2.6 ergab eine akzeptable Übereinstimmung für Säuren, aber eine größere Streuung für Basen. Die Studie kam zu dem Schluss, dass konsistente p*K*a-Werte, die mit mindestens zwei unabhängigen Methoden ermittelt wurden, die OECD-Kriterien für die Verwendung in der Regulierung erfüllen.

Insgesamt wurde bestätigt, dass NAMs, QSARs und verfeinerte physikalisch-chemische Parameter gemeinsam zuverlässige Informationen liefern können, um unter definierten Bedingungen Bedenken hinsichtlich der Bioakkumulation von Stoffen auszuschließen. Ihre Anwendung ist jedoch für superhydrophobe, ionisierbare und nicht vorwiegend in Lipiden akkumulierende Stoffe eingeschränkt, was fallspezifische Anpassungen oder ergänzende Tests erforderlich machen kann.

Bewertung der integrierten Test- und Bewertungsstrategie

Die Bewertung der integrierten Bewertungsstrategie konzentrierte sich auf die Untersuchung der Kohärenz, Zuverlässigkeit und Vorhersagekraft der kombinierten Methoden für 26 Fallstudienstoffe. Die Analyse verglich Datensätze aus dem *Hyalella azteca*-Biokonzentrationstest (OECD TG 321), In-vitro-Assays (OECD TG 319A/B), In-vivo-BCF-Studien an Fischen (OECD TG 305) und QSAR-Vorhersagen. Ziel war es, festzustellen, ob der integrierte, mehrstufige Ansatz eine wissenschaftlich fundierte Grundlage für regulatorische Entscheidungen bietet und gleichzeitig die Notwendigkeit von Tests an Wirbeltieren verringert.

Bei Substanzen mit geringer Hydrophobie reichten physikalisch-chemische Screenings und QSAR-Schätzungen aus, um eine Nicht-Bioakkumulation festzustellen, sodass theoretisch keine weiteren Tests erforderlich waren. Bei mäßig bis stark hydrophoben Substanzen traten jedoch bei bestimmten Chemikalien Abweichungen zwischen HYBIT, in vitro und In-vivo-Daten auf, was die Bedeutung einer fachkundigen Interpretation der Ergebnisse unterstreicht. Bei stark hydrophoben Substanzen war die Nutzung der aktuell verfügbaren NAMs aufgrund methodischer Schwierigkeiten eingeschränkt. Die Studie hebt hervor, dass für diese Stoffe gezielte Strategien und weiterführende Forschung erforderlich sind.

Die Bewertung befasste sich auch mit Stoffen, die sich nicht vorwiegend in Lipidphasen anreichern, darunter ionisierbare Verbindungen, PFAS und Tenside. Ionisierbare Stoffe wurden anhand von log D- und p*K*a-abhängigen Verteilungsmodellen bewertet, während PFAS und Tenside eine spezifische Berücksichtigung des Bindungs- und Verteilungverhaltens erforderten. Die derzeitigen Testmethoden lieferten erste Erkenntnisse, waren jedoch nicht universell anwendbar, was die Notwendigkeit einer weiteren methodischen Entwicklung unterstreicht.

Für luftatmende Organismen zeigten In-vitro-Biotransformationsdaten aus Rattenhepatozyten-Assays Potenzial, aber die Abschätzung des Bioakkumulationspotenzials durch IVIVE-Modelle muss noch weiterentwickelt werden. Obwohl sich dieser Ansatz noch in einem frühen Stadium der Validierung befindet, zeigte die Verwendung von Ratten-In-vitro-Assays jedoch ein starkes Potenzial für die Bewertung der Biomagnifikation ohne Tierversuche. Der vorgeschlagene Schwellenwert der Stufe 2, bei dem eine Abbaukonstante von mehr als 3 h^{-1} in Ratten-S9-Assays auf eine geringe Bioakkumulation hinweist, erwies sich als praktisches Screening-Kriterium, muss jedoch noch weiter verfeinert und an Rattenhepatozyten angepasst werden.

Zusammenfassend hat sich die integrierte Test- und Bewertungsstrategie als wirksam erwiesen, um Stoffe mit geringem Bioakkumulationspotenzial zu identifizieren und effiziente Testabläufe zu steuern. Sie bestätigte die komplementäre Rolle von In-silico-, In-vitro- und Wirbellosen-Methoden und unterstrich die Bedeutung von Expertenurteilen in komplexen Fällen. Mit Hilfe von NAMs lassen sich Tests an Wirbeltieren erheblich reduzieren, ohne dass die regulatorischen Schutzniveaus beeinträchtigt werden.

Abschließende Überlegungen und zukünftige Wege

Das in diesem Projekt vorgestellte integrierte Bewertungssystem bietet einen umfassenden und zukunftsorientierten Rahmen für die Bewertung der Bioakkumulation in verschiedenen Chemikalienvorschriften. Es ermöglicht den Regulierungsbehörden, sich von der veralteten Abhängigkeit von Tests an Wirbeltieren zu lösen, unterstützt die Bewertung einer Vielzahl chemischer Substanzen und fördert die Kohärenz zwischen den verschiedenen Vorschriften.

Das integrierte Bewertungssystem basiert auf einer flexiblen fünfstufigen Struktur und legt den Schwerpunkt auf frühzeitiges Screening, die Verwendung von In-silico-Daten und NAMs, um Tierversuche auf ein Minimum zu reduzieren. Die Studie zeigt, dass NAMs unter bestimmten Bedingungen Bedenken hinsichtlich der Bioakkumulation wirksam ausschließen können, wodurch die Abhängigkeit von Tierversuchen verringert wird. Bei problematischen Stoffen sind jedoch weiterhin kritische Überprüfungen und ergänzende Methoden unerlässlich.

Vor dem Hintergrund wachsender Umweltbelastungen, technologischer Möglichkeiten und gesellschaftlicher Erwartungen bietet das integrierte Bewertungssystem ein Konzept für die nächste Generation der Chemikaliensicherheit – ein Konzept, das evidenzbasiert, effizient und ethisch fundiert ist. Seine Einführung wird nicht nur die wissenschaftliche Qualität regulatorischer Entscheidungen verbessern, sondern auch zu einem nachhaltigeren und schützenderen Ansatz für die Umweltgesundheit beitragen.

1 Introduction

The evaluation of bioaccumulation potential is a cornerstone of environmental risk assessment and regulatory chemical evaluation. From the moment it became evident that certain substances can accumulate in organisms to harmful levels, bioaccumulation has played a vital role in chemical regulation. This significance is rooted in both scientific insight and growing societal awareness of environmental protection, sparked notably by Rachel Carson's influential work *Silent Spring* (Carson, 1962). Her book catalysed environmental consciousness globally and led to the development of broad environmental legislation across jurisdictions.

Initially, bioaccumulation assessment of chemicals focused on aquatic organisms, particularly fish. This led to the development of the OECD Test Guideline (TG) 305, originally published as a series (TG 305 A–E) and later revised into a single harmonised version in 1996, with a further update in 2012 (OECD 2012). For decades, this fish bioconcentration test has served as the primary tool for assessing the bioaccumulation potential of chemical substances. However, regulatory science has evolved significantly. Bioaccumulation is now part of the criteria for Persistent Organic Pollutants (POPs) under the Stockholm Convention (UNEP 2001) and for Persistent, Bioaccumulative and Toxic (PBT) substances under REACH (EC 2006) and other EU chemical frameworks. Moreover, advances in chemical production have brought forth new challenges, such as the emergence of PFAS, siloxanes and others - substances that do not always fit into existing regulatory models. Importantly, research has demonstrated that the bioaccumulation behaviour of chemicals in air-breathing organisms may differ from that in aquatic species (Schwarz & Gobas 2003). This has highlighted a gap in current assessment strategies and the need for a more comprehensive evaluation framework. In parallel, the demand to reduce vertebrate testing - consistent with the 3R principles of replacement, reduction, and refinement - has become a regulatory imperative (Russel & Burch 1959, Balls & Combes 2005, National Research Council 2007, European Parliament and Council 2010). This shift has stimulated the development of non-animal alternative methods, including in vitro systems and tests using invertebrates. Responding to this shift, new test guidelines such as OECD TG 319 (in vitro biotransformation) and OECD TG 321 (*Hyalella azteca* bioconcentration test) have been developed (OECD 2018a, OECD 2018b, OECD 2024a). These advancements have been acknowledged in the revised REACH Guidance R.11 (ECHA 2023) which begins to incorporate such alternative methods. Nevertheless, the practical integration of these new tools into regulatory decision-making remains limited and fragmented. From a regulatory perspective, there is still a lack of experience with the new test methods and a lack of confidence in their suitability for regulatory purposes.

Also, bioaccumulation assessment under REACH, as well as under biocidal, plant protection and medical products regulations differs, each apply different criteria and rely on distinct testing frameworks for PBT and secondary poisoning assessment. The majority of regulatory decision-making still depends heavily on fish-based bioconcentration factors (BCFs) for bioaccumulation assessment even though their applicability to more polar, ionisable, or structurally diverse compounds remains scientifically uncertain. Moreover, environmental monitoring studies have revealed substances that show minimal bioaccumulation in aquatic species but significant accumulation in terrestrial organisms, i.e. mammals (e.g. Kelly et al. 2007b). One example for this is PFOS, which has been listed as a POP in the Stockholm Convention due to its accumulation and long half-lives in mammals, although the fish BCF is < 5000.

Project Objectives

In light of these challenges, the German Environment Agency (UBA) has funded a comprehensive research project to contribute to modernising, optimising and harmonising bioaccumulation assessment within EU regulatory frameworks. A central focus of the project was the thorough evaluation of the opportunities, reliability, and limitations of both traditional and alternative test methods, forming the foundation for developing an integrated assessment strategy that meets regulatory needs while reducing vertebrate animal testing and building on existing strategies. To this end, the study centred on a comparative analysis of 26 substances for which OECD TG 305 fish bioaccumulation data were available, as well as either the *Hyalella azteca* test (OECD TG 321) and/or in vitro approaches (OECD TG 319 A/B; in vitro rat hepatocyte biotransformation assay). Where such data were missing, new experimental tests were conducted within the project. The selected substances covered a broad range of chemical properties, including compounds at the edges of the applicability domains for various test methods. The comparison of the various test results aimed at improving the understanding if a decision on the bioaccumulation of a chemical can be taken with sufficient confidence on the basis of the *Hyalella azteca* test (OECD TG 321) and/or in vitro approaches (OECD TG 319A/B), and to identify any constraints.

In addition to experimental testing, a literature review and database analysis was carried out. A list of 231 chemicals spanning pharmaceuticals, biocides, UV filters, plasticisers, POPs, pesticides, surfactants, and other high-concern chemicals—was compiled. QSAR-based consensus modelling of log K_{OW} values was performed, and additional physicochemical parameters such as log K_{OA} and pK_a were analysed. Estimation methods for deriving fish BCF values from QSAR models were also evaluated, highlighting methodological uncertainties and options for improvement. The outcome of this multi-layered assessment was a structured comparison of traditional and alternative test systems.

Ultimately, the project aims to support a paradigm shift in bioaccumulation assessment - away from reliance on single-method vertebrate testing toward a more flexible, tiered, and integrated approach. Therefore, the project proposes an integrated assessment strategy for bioaccumulation assessment across chemical regulations. By incorporating physicochemical screening, in silico, in vitro, and invertebrate test systems, and by aligning assessment schemes with the latest scientific evidence, the project contributes to the development of more ethical, efficient, and chemical regulation systems without lowering the level of protection. The decision schemes within regulatory frameworks could be optimised in various ways e.g. to reduce the use of vertebrates or to allow faster decision making with limited tiers. Since the data requirements in the different regulations differ, as well as the regulatory consequences of identification of a specific bioaccumulation potential, the regulatory implementation has to be left to the relevant regulatory committees.

A key milestone in this project was a workshop held in Berlin in March 2025, which brought together stakeholders from science, industries and European regulatory agencies. The workshop served as a forum for presenting the project's findings, discussing implications for regulatory practice, and identifying priorities for future harmonisation and method validation (see Annex A.13 for workshop summary).

To guide the development of an advanced bioaccumulation assessment strategy, this report progresses from regulatory background to detailed method analysis. Chapter 2 outlines current EU regulatory frameworks across various sectors, including REACH, biocides, plant protection and medicinal products. Chapter 3 introduces existing integrated testing strategies, while chapter 4 presents a proposed framework for assessing bioaccumulation in both aquatic and terrestrial organisms. Chapter 5 provides a summary of the results from an evaluation of

conventional and alternative test methods for which in vivo (OECD TG 321), in vitro (OECD TG319A/B, and rat S9 and hepatocyte biotransformation assays), and in silico approaches were carried out. A detailed description of the study outcome is provided in the Annex. Using data from the literature which were complemented by results generated in this project, the proposed integrated assessment strategy was evaluated (chapter 6) against this data. Chapter 7 summarises the main conclusions, identifies research gaps and makes recommendations for improving regulatory bioaccumulation assessment, with the ultimate goal of eliminating the need for animal testing while guaranteeing the highest level of public and animal health and environmental protection.

Structure of the report

The report comprises seven principal chapters forming the main part, followed by a comprehensive Annex. The Annex provides supporting experimental data, detailed methodological descriptions, and supplementary analyses, ensuring full transparency and traceability of results. The supplementary materials provide detailed datasets and supporting information used in the bioaccumulation assessment. They include raw data on key physicochemical parameters (log K_{OW}, log K_{OA}, and pKa), quantitative structure–activity relationship (QSAR) results for bioconcentration factors (BCF), and further substance-specific data endpoints.

1.1 Definitions and Disclaimers

Project objectives

Scope of the project

Evaluation of the state of NAMs: One aim of this project is to collect NAM data for a series of different chemicals and to compare the endpoints derived from different NAMs against fish data. This comparison shall provide regulators with a better picture of the current power of NAM endpoints for their decision-making.

Explore boundaries of NAMs: Regulatory acceptance of NAMs depends on an understanding of the methods' boundaries. Applicability domains but also limits of the NAMs are to be explored and explained.

Evaluate the demands of different European Legislations: The data requirements and implementation of bioaccumulation endpoints can differ among legislations. The specific demands of legislations will be evaluated against the applicability domains of the NAMs.

Integrated assessment strategy: All above mentioned aspects will be included in an integrated assessment strategy.

Out of scope

Binding regulatory guidance: The integrated assessment strategy is a proposal. The demands of different regulations and decision-making bodies can vary with regard to bioaccumulation endpoints. Primary protection goals may vary. Thus, integration into regulatory guidance is in the hands of the relevant bodies at ECHA, EFSA and EMA.

Detailed data quality analyses: The endpoints included in this report reflect the reality of decision makers: Not all data that is provided fulfils all criteria for a perfectly executed study.

NAMs

NAMs, or „New Approach Methodologies" is a term used to describe (test) methods, that aid the phasing out and ultimately the ban of the use of animal tests for the purpose of chemical regulation. A variety of test approaches can be included in this umbrella term.

"In silico methods" is an umbrella term for calculations of any complexity that are commonly performed using a computer.

"In vitro methods" refer to laboratory test systems that require biological material and are performed in small scale labware. The biological complexity of the biological material can range from protein extracts to cell organelles, whole cells, living organisms (bacteria, fungi), or even (parts of) organs. Exact definitions of "in vitro" may vary depending on ethical interpretations or national animal protection laws.

IVIVE is the abbreviation for "in vitro to in vivo extrapolation". IVIVE methods are used to transfer in vitro measurements to in vivo information. Usually, mathematical models are used to extrapolate and incorporate the in vitro data to in vivo scenarios.

"Animal tests" are tests that are performed using living animals. For the evaluation of bioaccumulation it is necessary to determine internal concentrations of the test chemical in the exposed organism. In laboratory tests, this commonly requires the sacrifice of the test animals. In monitoring studies, biopsies may provide enough sampling material, if the target animal is large enough. The term "animal" can be interpreted in a biological sense and in a regulatory context. The regulatory definition of "animal" may vary depending on the respective animal protection law. In Europe, "animal", for scientific purposes, refers to non-human vertebrates (*cf*. European Treaty Series - No. 123).

Understanding of NAMs in this report: NAMs in this report include test methods that are in silico, in vitro or in vivo methods that do not require the in vivo use of vertebrates.

Understanding of "animal" in this report: Animal is used in the context of the European Animal Welfare Law (European Parliament Council 2010)

Chemicals

The variety of chemicals is profound and has to be respected in regulatory assessment. The following descriptions detail key characteristics of chemicals that can have an influence on the way a chemical has to be assessed. A chemical can fulfil more than one of the following criteria.

Neutral, hydrophobic chemicals

The majority of work and development around bioaccumulation testing and assessment has revolved around neutral organic substances. Hence, most tests are designed to reliably assess their potential of bioaccumulation.

Superhydrophobic chemicals

Chemicals that show a high degree of hydrophobicity and can be harder to test: Their application in the water phase is more difficult since the chemical is more likely to adhere to organic compounds. This aspect can make these types of chemicals harder to assess. The exact definition of superhydrophobic chemicals is under debate at the current point of time. An ongoing UBA project (FKZ 3723 64 4050 "Alternative strategies for the assessment of bioaccumulation of superhydrophobic substances") is further exploring how to handle superhydrophobic chemicals in

regulatory terms. For the ease of reading, chemicals with a log K_{OW} of > 6 are referred to as superhydrophobic in this project, although this is not meant to be a strict cut-off: depending on the chemical and the respective test, testing may be feasible, or difficult, or not feasible.

Ionisable chemicals

Some chemicals can be ionised, depending on the pH of the medium the chemical resides in. A chemical may behave differently, if it is in an ionised or neutral state. Chapter A.12.1 provides a detailed explanation about the special characteristics of ionisable chemicals.

Non-lipid accumulating chemicals

OECD TG 305 requires a lipid normalisation to harmonise BCFs to a common lipid content. Lipids are not always the target compartment of chemicals. Protein/serum binding is another example for an interaction of specific chemicals and biological material. In such cases, lipid normalisation is not sensible.

Surfactants

Surfactants show a specific, amphiphilic behaviour. In aqueous solutions they are prone to sorb to surfaces or the water/air interface. At elevated concentrations they can form complexes known as micelles. This specific behaviour is a result of their structure: A polar head group that interacts well with water and a nonpolar tail that repels water. Surfactants can be neutral of charge, anionic, cationic and zwitterionic, depending on the molecule structure and pH of the surrounding medium. Hydrophobicity can vary depending on the tail length and structure. Log K_{OW} and log D are no suitable parameters to describe their partitioning behaviour. Details in chapter 4.1.2.

Organisms

Aquatic organisms

Environmental bioaccumulation assessment has been focused on aquatic organisms, i.e., fish. The fish bioconcentration factor is the ultimate decisive endpoint for the bioaccumulation criterion in European and other substance legislations. Decades of laboratory data collection allow for the development of in silico methods. To reduce the number of fish, i.e., vertebrates for testing, test methods applying in vitro and invertebrate approaches have been developed. For regulatory acceptance, their respective scope of application needs to be understood more clearly.

Terrestrial, air breathing organisms

Bioaccumulation data of air-breathing, terrestrial vertebrates has thus far been considered in form of monitoring data. Substances that bioaccumulate or biomagnify in air-breathing vertebrates, but not as much in aquatic communities, have been identified. To respect the precautionary principle of chemical regulation, air-breathing vertebrates must also be put in focus of chemical regulation prior to their market entry. This inclusion should not result in additional vertebrate testing.

Organisms with respiration via the skin

Organisms that regulate their air exchange via the skin, such as e.g., annelids, are represented in OECD TG 317. Bioaccumulation of this type of organisms are required under specific, regulatory cases but are not covered in this project.

Reporting

Artificial intelligence tools (DeepL, FhGenie, ChatGPT, Microsoft Copilot) were used to support the preparation of this report (e.g., translating, summarising, providing style suggestions, and spell checking).

2 Regulatory bioaccumulation assessment

Strategies for bioaccumulation assessment according to established European regulatory frameworks differ with respect to the testing methods applied, specific regulatory criteria, the promotion of alternative methods to minimise animal testing as well as the integration of comprehensive assessments.

2.1 REACH

The Regulation on the registration, evaluation, authorisation and restriction of chemicals (REACH) is one of the key EU legislative frameworks to protect human health and the environment from the risks that can be posed by chemicals (EC 2006) alongside the other sector-specific regulations. The PBT (Persistent, Bioaccumulative, and Toxic) assessment under the REACH regulation involves a systematic evaluation of a substance's properties based on specific criteria defined in Annex XIII of the regulation (EC 2006). If the log K_{OW} value of a chemical exceeds the trigger value of 4.5, a bioaccumulation assessment for aquatic organisms is required for the PBT/vPvB assessment. Additionally, a log K_{OW} greater than 2 together with a log K_{OA} greater than 5 serve as screening criteria for bioaccumulation assessment in air-breathing organisms (ECHA 2023a).

Bioaccumulation is essentially defined on the basis of the bioconcentration factor (BCF). A substance is considered to fulfil the criterion for “bioaccumulative” if the BCF in aquatic species is higher than 2000 (B criterion) indicating that the substance tends to accumulate in the tissues of organisms at a rate that exceeds the rate at which it is eliminated. According to the REACH regulation, a substance fulfils the criterion for “very bioaccumulative” (vB) substances, when the bioconcentration factor in aquatic species is higher than 5000. Apart from the BCF, other information, such as bioaccumulation in terrestrial species or monitoring data, may also be used in a WoE approach to assess the bioaccumulation potential, as specified in Annex XIII of the REACH regulation.

In most cases where experimental information on bioaccumulation in aquatic species is needed, an aqueous exposure bioconcentration fish test according to OECD TG305-I (or OECD TG 305-II) has been requested to obtain results for comparison with the B/vB criteria.

In October 2024, the REACH Member State Committee (MSC) unanimously approved the non-vertebrate *Hyalella azteca* test, HYBIT, to be used as the standard information requirement under REACH where applicable (https://echa.europa.eu/de/-/highlights-from-october-msc-meeting): “This recently adopted OECD Test Guideline 321 (OECD 2024a) assesses the bioaccumulation potential of substances in aquatic environments. It is based on non-vertebrate aquatic organisms and replaces testing on fish. Using the test as default is another step towards reducing testing on vertebrate animals”.

According to the revised ECHA Guidance on Information Requirements and Chemical Safety Assessment (ECHA 2023a), *H. azteca* BCF results converted to 3% lipid (BCF_{KL}, 3%) are preferred for a comparison against the PBT criteria according to REACH Annex XIII for B and vB properties, and deviations should be justified. A default lipid content of 3% (w/w) is used for normalizing BCF values which represents a fat content which is common for laboratory-raised amphipods. If a substance has a valid and plausible *H. azteca* BCF (3% lipid, w/w) > 2000 (indicating a significant accumulation in the test organism), the substance is defined as ‘B’, if the BCF exceeds 5000, it is considered ‘vB’. A *H. azteca* BCF (3%, w/w) < 1200 indicates ‘not B’ because even with a lipid normalisation to 5% (w/w) as applied for fish, the threshold value of 2000 would not be passed. Accordingly, a BCF < 3000, normalised to 3% lipid, indicates not vB’ for the aquatic compartment (ECHA 2023a; Table 1).

However, a conclusion of 'not B' or 'not vB' for the aquatic compartment can only be drawn if there is no other information indicating the contrary. For *H. azteca* BCF values, normalised to 3% lipid, between 1200 and 2000 and between 3000 and 5000, respectively, it cannot be excluded that due to the lower lipid content of the amphipods (3%, w/w) the bioaccumulation potential of a substance may be underestimated compared to fish (5%, w/w). If no conclusion can be made with the available data, and uncertainties remain about the bioaccumulation potential in aquatic organisms, further testing with fish should be considered as a last resort, if justified (ECHA 2023a).

Table 1: Threshold values to characterise the bioaccumulation potential, according to REACH guidance R11. (ECHA 2023a).

B classification	BCF fish (OECD 305) 5% lipid (w/w)	HYBIT BCF (OECD 321) 3% lipid (w/w)
not B	< 2000	<1200: not B 1200 – 2000: further test
B	2000 - 5000	2000 – 3000: B 3000 – 5000: further test
vB	> 5000	> 5000: vB

Note: For a robust conclusion, any other available information needs to be considered as part of a WoE approach.

2.2 Biocides

The Biocidal Products Regulation (BPR, Regulation (EU) 528/2012) concerns the placing on the market and use of biocidal products, which are used to protect humans, animals, materials or articles against harmful organisms like pests or bacteria, by the action of the active substances contained in the biocidal product. This regulation aims to improve the functioning of the biocidal products market in the EU, while ensuring a high level of protection for humans and the environment. All biocidal active substances have to undergo a formal PBT assessment. Annex XIII to the REACH Regulation sets criteria for substances that are persistent, bioaccumulative and toxic (PBT) or very persistent and very bioaccumulative (vPvB) as described in chapter 2.1. The BPR promotes the reduction of animal testing by encouraging the use of alternative testing methods. References to REACH guidance documents are made.

2.3 Plant protection products

The bioaccumulation potential of products, specifically active substances, safeners, or synergists in plant protection products (PPPs), is evaluated through criteria established in Regulation (EC) No 1107/2009 (EC 2009). An active substance is considered to have bioaccumulation potential if its bioconcentration factor in aquatic species is greater than 2000 and fulfils the 'very bioaccumulative' criterion where the bioconcentration factor is greater than 5000. In the absence of such data, a partition coefficient n-octanol/water (log K_{OW}) greater than 5, or evidence that the active substance, safener or synergist shows a high bioaccumulation in other non-target species can be used to assess the bioaccumulation criterion. No reference to REACH Annex XIII or the PBT guidance under REACH is made. Assessment of bioaccumulation shall be based on measured data on bioconcentration in aquatic species (Regulation (EC) No 1107/2009 (Annex II, chapter 3.7.2.2 – Bioaccumulation)). Data from both freshwater and marine water species can be used. However, fish flow-through tests according to OECD Test Guideline 305 are required according to the EC data requirements for testing the bioaccumulation potential of plant protection products in the regulatory context. Guidance for the bioaccumulation

assessment of PPPs is given by EFSA (EFSA 2013) and European Commission (EC 2012). The use of alternative methods to the fish test is not yet provided for in the regulatory framework.

2.4 Medicinal products for human use

The guideline on the environmental risk assessment of medicinal products for human use (HMPs) is based on Directive 2001/83/EC (EC 2001) and EMA ERA Guideline EMEA/CHMP/SWP/4447/00 Rev. 1, 2024. (EMA 2024). This directive pertains to the risks to the environment arising from the use, storage, and disposal of medicinal products, rather than risks from their synthesis or manufacture. The guideline specifically focuses on environmental risks associated with the use of human medicinal products.

A bioconcentration factor (BCF) in fish should be determined experimentally in accordance with established guidelines (such as OECD TG 305) when the log K_{OW} for the active substance is ≥ 3. If the BCF is ≥ 100 L kg^{-1}, the potential for secondary poisoning should be further assessed. If the BCF is > 2000 L kg^{-1}, additional evaluations for PBT/vPvB classification may be warranted. If the log K_{OW} value exceeds the trigger value of 4.5, a definitive PBT/vPvB assessment is required. The tiered approach allows for a systematic evaluation, starting from the screening level and progressing to more detailed assessments if the initial criteria are met. For dissociating compounds, the BCF test should be conducted at a stable, environmentally pH-range from 5 to 9 that is consistent with the most bioaccumulative form of the test chemical. This is typically the non-dissociated form or the form with the most neutral molecule species (EMA 2024). ^{14}C radiolabelled testing is not a default requirement for human medicinal product ERA under the EMA guideline, but it is a common and sometimes essential analytical tool in bioaccumulation or degradation studies when standard analytical methods lack the sensitivity or specificity to follow the substance and its transformation products.

The guideline primarily focuses on aquatic bioaccumulation studies, specifically referencing the OECD TG 305 guideline for assessing bioaccumulation in fish. It emphasises that bioaccumulation is typically evaluated in aquatic species rather than terrestrial organisms. The use of alternative methods to the fish test is not yet provided for in the regulatory framework, but also not excluded based on R.11 (ECHA 2023a), to which the regulation refers.

2.5 Veterinary medicinal products (VMPs)

The requirement for assessment of environmental safety for Veterinary Medicinal Products (VMPs) is regulated by the German Veterinary Medicines Act, which implements EU Regulation 2019/6 (EU 2019). Details on environmental risk assessment are provided in VICH (Veterinary International Conference on Harmonisation) guidelines GL 6 and GL38 (CVMP/VICH Guidelines 592/98 (GL 6), 790/03 (GL 38), EMA/CVMP/ERA/418282/2005-Rev.1- Corr.) (EMA 2016).

Steps for assessing the bioaccumulation potential of VMPs are the determination of physicochemical properties such as the octanol-water partition coefficient (log K_{OW}), which helps to predict the bioaccumulation potential. A log K_{OW} value greater than 4 is a key indicator of potential bioaccumulation and necessitates careful consideration in environmental risk assessments for VMPs. If the log K_{OW} indicates a potential for bioaccumulation, BCF studies may be required. In the context of environmental impact assessments for VMPs, the OECD TG 305 is commonly used to evaluate the risks associated with their potential accumulation in aquatic environments. In addition, Absorption, Distribution, Metabolism, and Excretion (ADME) studies may help to understand how the VMP is metabolised in organisms and to identify whether metabolites also have bioaccumulation potential. Some metabolites may be more hydrophilic than the parent compound and thus be less likely to bioaccumulate.

The assessment of a terrestrial risk becomes necessary when log K_{OW}/D_{OW} and specific triggers for a probability of occurrence in the soil are exceeded. If a potential risk for poisoning in the terrestrial food chain is identified, further testing with terrestrial organisms, e.g. a test on bioaccumulation in terrestrial oligochaetes (OECD TG 317) or feeding studies with laboratory mammals or birds may be required.

In case of lack of experimental data, especially for the assessment of terrestrial exposure, initial predictions based on QSARs can be used in the risk assessment (secondary poisoning via food chains). While in vivo fish testing remains a key component of bioaccumulation assessment, the EMA's "Regulatory Science to 2025" strategy — supported by the Committee for Medicinal Products for Veterinary Use (CVMP) — promotes the integration of innovative, alternative methods as a viable approach to environmental risk assessments of VMPs (European Medicines Agency, 2020). However, an integrated testing and assessment strategy providing a cohesive strategy for evaluating the safety and efficacy of substances without relying solely on traditional vertebrate testing is still missing in the regulatory framework.

2.6 Classification, Labelling and Packaging (CLP)

The Classification, Labelling and Packaging (CLP) Regulation ((EC) No 1272/2008) (EC 2008) is based on the United Nations' Globally Harmonised System (GHS) and its purpose is to ensure a high level of protection of health and the environment, as well as a free market for substances, mixtures and articles. CLP is legally binding across the Member States and directly applicable to all industrial sectors (EC 2008). Thus, it is also applicable to biocides and plant protection products. It requires manufacturers, importers or downstream users of substances or mixtures to classify, label and package their hazardous chemicals appropriately before placing them on the market (EC 2008). The guidance (ECHA 2024) refers closely to the REACH guidance. A BCF in fish of ≥ 500 is indicative of the potential to bioconcentrate for classification purposes. Some relationships can be observed between chronic toxicity and bioaccumulation potential, as toxicity is related to the body burden (ECHA 2024). Experimentally derived BCF values of high quality are preferred for classification purposes. BCF results from poor or questionable quality studies should not be used for classification purposes if high quality data on log K_{OW} are available. If no BCF is available for fish species, high quality data on the BCF for some invertebrates (e.g. blue mussel, oyster and/or scallop) may be used as a worst-case surrogate (ECHA 2024). BCF estimates based on in vitro OECD TGs 319A and B might be considered in a WoE approach for bioaccumulation provided that they fulfil relevant data quality requirements (ECHA 2024).

2.7 Secondary poisoning

Secondary poisoning (SP) may occur when an organism is exposed to a toxicant indirectly through the food chain, rather than by direct contact with the contaminated environment. In environmental regulation (e.g., under REACH in the EU), secondary poisoning is explicitly assessed when a substance has both bioaccumulation potential (often a log K_{OW} > 3) and toxicity to higher organisms. In this case, predicted oral exposure is compared to respective toxicity thresholds (e.g., avian or mammalian NOAELs) to determine ecological risks. This evaluation typically uses experimental fish BCFs and BMFs (biomagnification factors) or BSAFs (Biota-to-Soil Accumulation Factors) from studies on terrestrial oligochaetes (OECD 2010) to determine how substances accumulate in prey organisms (e.g., fish, earthworms) and how they are subsequently transferred to their predators in the food chain (e.g. ECHA 2016a). In addition to REACH, secondary poisoning assessment is also addressed in the EU regulatory frameworks for biocidal products, plant protection products, veterinary medicinal products, and human medicinal products (Table 2). For example, for plant protection products, the assessment of SP is

described in the GD on "Risk Assessment for Birds and Mammals" (EFSA 2009), assessing the accumulation of the chemical via consumption of earthworms into small birds and mammals or of fish into top predators.

Table 2: Assessment of secondary poisoning based on bioaccumulation data.

Regulation	Trigger for Secondary Poisoning Assessment	Typical Data / Metrics	Relevant Test Methods	Main Target Predators Considered
REACH (Reg. (EC) No 1907/2006); Guidance on Information Requirements and Chemical Safety Assessment, Chapter R.7c (ECHA 2023b)	Substance has bioaccumulation potential (e.g., high log K_{OW}, BCF > 2000, BMF > 1, BSAF > 1) and mammalian/avian toxicity	Log K_{OW}, BCF (fish), BMF, BSAF (earthworms), NOAELs	OECD TG 305 (fish bioaccumulation), OECD TG 317 (earthworm bioaccumulation)	Aquatic (fish-eating mammals/birds), terrestrial (worm-eating mammals/birds)
Biocidal Products (Reg. (EU) No 528/2012); Guidance on the Biocidal Products Regulation: Volume IV: Environment - Part A (ECHA 2022a) and Part B+C (ECHA 2017a)	Bioaccumulation potential + toxicity; high risk for non-target species likely	BCF, BMF, residue levels in target and non-target organisms	OECD TG 305, TG 317; field monitoring data	Predatory birds/mammals (esp. for rodenticides, insecticides)
Plant Protection Products (Reg. (EC) No 1107/2009); Guidance document (EFSA 2009)	Potential for residues in prey species and risk to higher trophic levels	Residue studies in invertebrates/vertebrates, BCF/BMF	OECD TG 305, TG 317; field residue trials	Earthworm (BCF calculated or experimentally derived)
Veterinary Medicinal Products (Reg. (EU) 2019/6), CVMP/VICH Guidelines 592/98 (GL 6) and 790/03 (GL 38), EMA/CVMP/ERA/418282/2005-Rev.1- Corr.1	Log K_{OW} ≥ 4; BCF ≥ 1000 Evidence from metabolism/residues/excretion, biodegradation studies and molecular mass should be considered	Log K_{OW}, BCF aquatic, terrestrial BMF (default values if no experimental data is available) Mammalian or bird toxicity data (NOAELs, NOECs)	OECD TG 305, TG 317* toxicity studies reporting on dietary and oral exposure	Aquatic and terrestrial food web: Birds (e.g. vultures); mammals
Human Medicinal Products (Dir. 2001/83/EC & EMA ERA Guideline EMEA/CHMP/SWP/4447/00 Rev. 1, 2024)	Log K_{OW}/Dow ≥ 3; BCF≥100	Log K_{OW}/Dow, BCF BMF if applicable, BSAF if relevant Mammalian or bird toxicity data (NOAELs, NOACs)	OECD TG 305, TG 317* if relevant; Non-clinical data (repeated dose toxicity, oral administration)	Predators in aquatic/terrestrial food webs (fish, birds, mammals)

* No standard requirement

2.8 Conclusion on current regulatory demands

The comparison of the different EU regulatory frameworks with regard to their test and assessment strategies for bioaccumulation assessment highlights the following key points:

- The bioconcentration factor (BCF) in aquatic organisms is the primary criterion for assessing the bioaccumulation potential of chemicals.
- The fish bioconcentration test according to OECD TG 305 still serves as the primary tool for assessing the bioaccumulation potential of chemical substances across different regulatory frameworks.
- The screening criteria (log K_{OW}-thresholds) and standard information requirements for bioaccumulation and secondary poisoning assessment vary across regulations. The possibility of using alternative approaches in a WoE, if no direct or conclusive data are available, differs considerably.
- New Approach Methodologies, such as OECD TG 319 and OECD TG 321, are available, but the practical integration of these new tools into EU regulatory bioaccumulation assessment remains limited.
- The recent approval of the *Hyalella azteca* bioconcentration test (OECD TG 321) as a standard information requirement under REACH is significant. It allows for assessing bioaccumulation potential using non-vertebrate organisms, thereby reducing reliance on in vivo fish testing, which aligns with the EU's goal of minimising, and in the long-term phasing out, animal testing.
- The assessment of bioaccumulation is a multifaceted process, incorporating experimental data, theoretical models (like QSAR), and an understanding of the physicochemical properties of substances. Only a comprehensive approach for bioaccumulation assessment ensures a thorough risk evaluation for both environmental safety and human health.
- Regarding the different EU regulatory frameworks, the development of more consolidated testing and assessment approaches that integrate NAMs but leave sufficient options for the various regulatory demands, without lowering the protection level is required.

3 Review of existing integrated testing strategies for bioaccumulation assessment

Several Integrated Testing Strategies (ITS) for evaluating the bioaccumulation potential of organic chemicals have been presented. The ITS for bioconcentration of organic chemicals in fish developed by de Wolf et al. (2007) is a 4-tier approach which includes physicochemical analyses; computer models to estimate ADME properties and BCF (tier 1), in vitro assays to evaluate ADME properties (tier 2), modified in vivo methods to measure bioconcentration (tier 3), and fish flow-through test according to OECD TG 305 as "gold standard" (tier 4). The authors conclude that there are many opportunities for applying the principles of the 3Rs: Reduce, Refine, and Replace when the use of fish is addressed for the bioaccumulation assessment of chemical products and encourage the development of validated test guidelines on NAMs which took place over the last years.

The purpose of the ITS for bioaccumulation assessment presented by Lombardo et al. (2014) was to provide a systematic approach to evaluate the bioaccumulation potential of organic chemicals in alignment with REACH regulations. Waiving schemes are employed as well as analysis of physicochemical properties relevant to bioaccumulation. The strategy encourages the use of alternative methods, both in silico and in vitro, to minimise the need for in vivo (animal) testing. In vivo methods are recommended only as a last resort, thus significantly reducing the number of animal tests required. By limiting the number of in vivo tests needed, the ITS would contribute to a more cost-effective and timely chemical hazard assessment process. The proposed ITS was found to allow for waiving in vivo bioaccumulation testing for about 67% of tested substances, thereby promoting a more ethical approach to chemical safety assessment (Lombardo et al. 2014).

In 2022, a discussion paper on the "bioaccumulation assessment of air-breathing mammals" was shared by the European Chemicals Agency (ECHA) which was developed by the ECHA working group on toxicokinetics (ECHA 2022). The aim of the working group was to develop a formal assessment scheme to evaluate chemical bioaccumulation in air-breathing organisms, particularly mammals, intended to support chemical assessments such as those required under the REACH regulation. The document highlights the importance of integrating multiple lines of evidence, including data from in silico, in vitro, and in vivo studies, to support a comprehensive assessment. The working group calls for the establishment of a WoE approach to improve the reliability and relevance of assessments, allowing for better decision-making regarding the bioaccumulation potential of various chemicals. It acknowledges ongoing challenges in the assessment of bioaccumulation in air-breathing species and envisions the development of an assessment tool that aligns with OECD guidelines to facilitate the evaluation process.

The Bioaccumulation Assessment Tool (BAT) is a spreadsheet-based tool to guide the collection, generation, evaluation, and integration of various lines of evidence for aquatic and air-breathing (mammalian) organisms (Arnot et al. 2023). BAT Ver.1.0 was publicly released in October 2018 following its development with input from a multistakeholder community. It incorporates a quantitative WoE approach, which critically evaluates data reliability and integrates diverse data sources to inform decision-making related to bioaccumulation assessments. The tool utilises different lines of evidence, including in vivo, in vitro, and in silico methods, facilitating the evaluation of bioaccumulation potential through various metrics such as bioconcentration factor (BCF), biomagnification factor (BMF), and bioaccumulation factor (BAF). BAT is implemented in Visual Basic for Applications (VBA) with a user-friendly graphical interface in Excel™, making it accessible for users to navigate and apply in assessments. The tool is available for free download (https://arnotresearch.com/bat/) ensuring that stakeholders from academia, government, and

industry can utilise it for environmental toxicology assessments. The BAT, accompanied by guiding principles in the BAT manual (Armitage et al., 2021), promotes standardised recording and evaluation of various lines of evidence related to the endpoint bioaccumulation, but its use is time-consuming and the result not always helpful in a regulatory context. While the BAT standardises evidence evaluation, a key limitation is that, in regulatory practice, "reliable evidence of bioaccumulation cannot be outweighed by information showing no bioaccumulation" (ECHA 2023a).

The Integrated Approaches for Testing and Assessment (IATA) for bioaccumulation published by OECD (OECD 2024c) is a structured and iterative framework designed to evaluate the bioaccumulation potential of chemicals through a weight-of-evidence (WoE) approach that integrates multiple lines of evidence (LoE). The process follows a tiered testing strategy that progresses from less to more resource-intensive methods, enabling refinement and the closure of data gaps as needed. It combines diverse data sources, including in silico models, in vitro assays, and in vivo studies, which are evaluated for their relevance and reliability. The strength of evidence (SoE) is assessed by examining the coherence and consistency among data sets and assigning weights based on reliability and relevance. This helps to quantify uncertainty and increase confidence in the overall evaluation. Uncertainties arising from data gaps or inconsistencies are identified, categorised qualitatively, and addressed through targeted data generation when necessary. In the final stage, all evidence is integrated to assess coherence, remaining uncertainties, and potential data gaps, ensuring transparent reporting to support regulatory or scientific decision-making.

The main difference between established regulatory assessment schemes and the new OECD IATA for bioaccumulation lies in their methodology, flexibility, and approach to evidence evaluation. Established regulatory assessment schemes (see chapter 2) typically rely on single, resource-intensive testing methods, such as standard in vivo bioaccumulation studies, to support regulatory decisions. These approaches usually follow prescriptive and rigid testing protocols that emphasise individual test outcomes rather than integrating multiple lines of evidence in a systematic manner. As a result, they can be inefficient and time-consuming, often requiring extensive animal testing and incurring high costs. Moreover, the established schemes generally lack explicit procedures for addressing uncertainty or for systematically integrating and weighting different types of data, which can limit the transparency and robustness of the final assessment.

Generally, there is a need for a scientifically sound and animal-sparing strategy for assessing the bioaccumulation potential of substances across all regulatory systems. However, differences in threshold criteria and interpretive frameworks between jurisdictions create barriers to harmonisation and regulatory acceptance of comprehensive assessment schemes such as the new OECD IATA. The overarching objective is thus to evolve the existing assessment frameworks in a way that maintains the same level of protection as current EU regulatory approaches, while fostering innovation and reducing reliance on animal testing. The refined approaches should incorporate alternative methods (e.g., in silico, in vitro, and invertebrate tests) and be applicable to substances with challenging properties such as ionisability, high hydrophobicity, or volatility. They should also account for other modes of bioaccumulation than lipid accumulation, such as sorption (e.g., PFAS) or amphiphilic distribution (e.g., surfactants).

The proposed assessment system (chapter 4) provides the basis for the further development of the established regulatory assessment frameworks with the aim of reducing animal testing on vertebrates to a necessary minimum by taking into account in silico methods, in vitro methods or tests on invertebrates.

4 Proposal for an integrated assessment strategy for aquatic and terrestrial bioaccumulation

Based on the existing integrated testing strategies, the guidance for the European chemical substance regulations (ECHA 2023a, b) and the test methods in the scope of the project (see chapter 5), an integrated assessment scheme for evaluating the bioaccumulation potential of substances of different substance classes in aquatic organisms was developed, evaluated and assessed (see chapter 6). The focus was on laboratory testing for aquatic bioconcentration, with the fish bioconcentration test (OECD 305 I) and the NAMs with the same endpoint, which are the HYBIT (OECD TG 321) and the in vitro tests (OECD TG 319A/B). Also, the rat in vitro test was included to account for bioaccumulation in air-breathing organisms. Monitoring data and additional information from further studies are also considered, if available, as part of a WoE approach.

4.1 Integrated assessment strategy for aquatic bioaccumulation

The tiered assessment approach for evaluating bioaccumulation potential of organic chemicals in aquatic, water-respiring organisms is presented in Figure 1. Specific thresholds for assessing bioaccumulation at the various assessment levels (tiers 1-4) need to be determined in accordance with the assessment criteria of the respective regulatory frameworks. The threshold values enable the bioaccumulation potential of substances to be assessed or provide indication that further testing at the next higher, more complex level of the assessment scheme is necessary. The protective effect of these values and conditions should be at least comparable with the effect of the current regulatory assessment concepts (see chapter 2). The assessment of secondary poisoning based on bioaccumulation data is also part of the proposed assessment scheme. The scheme is not only applicable to the "classic" neutral, hydrophobic substances. Substances that are, for example, ionisable, or superhydrophobic are also considered. Recommendations for the bioaccumulation assessment of compounds that follow different modes of bioaccumulation such as non-lipid accumulation (e.g. PFAS) or amphiphilic distribution (e.g. surfactants) are provided.

In the assessment of chemicals, two scenarios can be distinguished:

For a new chemical, no information on bioaccumulation is available, thus the assessment strategy is meant to give guidance for testing and how to proceed to produce the information needed for decision making. In a regulatory context, the data requirements of the respective regulation also have to be met. For other substances, a collection of data may already be available. In this case, all available, relevant and reliable information (e.g. monitoring data and additional information on bioaccumulation potential) should always be taken into account in addition to data submitted in the registration/authorisation process in order to conclude on the property in a WoE assessment, even if lower tier information suggests that the assessment might be concluded at that stage.

4.1.1 Bioaccumulation predominantly in lipid

The assessment approach for evaluating bioaccumulation potential of organic chemicals in aquatic, water-respiring organisms encompasses five tiers. Tier 1: In a first step, a screening based on the chemicals' hydrophobicity, as the key physicochemical property, is conducted. The screening also includes checking for ionisable or amphiphilic properties or others. Information on the testing of chemicals showing non-lipid accumulation is presented in chapter 4.1.2.

The log K_{OW} is commonly used as a measure of hydrophobicity. It can be determined experimentally or can be estimated by different QSARs. We suggest using the consensus modelling approach as detailed in chapter 5.4 to increase confidence. In case of ionisable chemicals, the log D for the realistic worst-case pH should be used instead.

The specific threshold values above which bioaccumulation is considered potentially relevant differs between the various regulations, as described in chapters 2.1 - 2.7.

REACH: log K_{OW} > 4.5: Bioaccumulation assessment for aquatic organisms is required for PBT/vPvB assessment, as well as for secondary poisoning. Additionally, log K_{OW} > 2 together with log K_{OA} > 5: Screening criteria for bioaccumulation assessment in air-breathing organisms.

Biocides: log K_{OW} ≥ 3: bioaccumulation assessment for aquatic organisms is required as well as assessment of secondary poisoning; with regard to PBT/vPvB assessment the same trigger as for REACH is applied (log K_{OW} > 4.5)

Plant protection products: log K_{OW} > 3: Bioaccumulation assessment for aquatic organisms, for PBT/vPvB assessment, as well as for secondary poisoning.

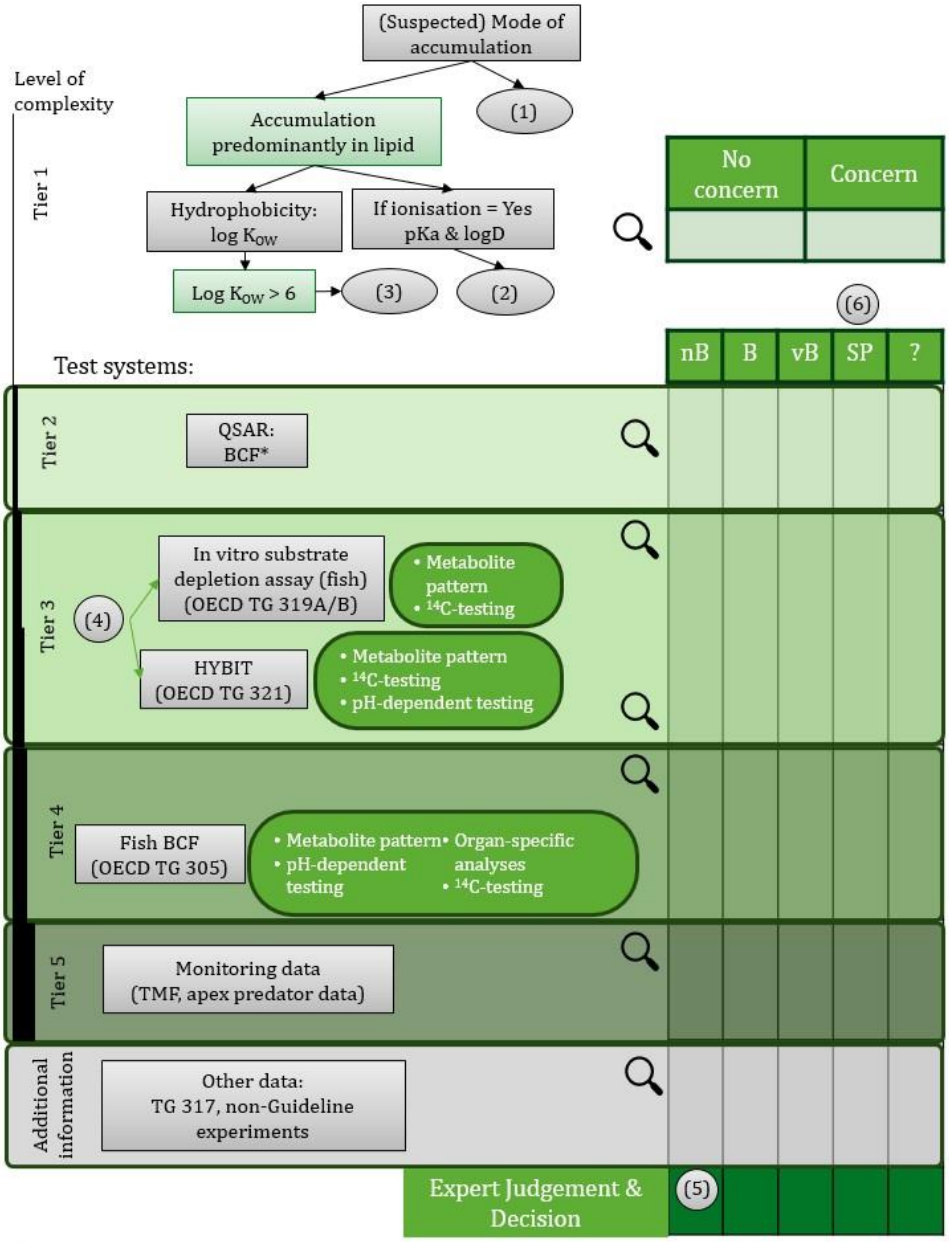
Medicinal products for human use: log K_{OW} ≥ 3: Bioaccumulation assessment for aquatic organisms is required. If the log K_{OW} value exceeds the trigger value of 4.5, a definitive PBT/vPvB assessment is required.

Veterinary medicinal products: log K_{OW} > 4: bioaccumulation assessment for aquatic and terrestrial organisms is required.

Thus, if

- the log K_{OW} < 2: Low potential to bioaccumulate or biomagnify in aquatic or terrestrial food chains, in air-breathing wildlife or in humans.
- log K_{OW} ≥ 2: if also log K_{OA} is ≥ 5, there is a potential to biomagnify in terrestrial food chains, in air-breathing wildlife and/or in humans. If no bioaccumulation in aquatic organisms is found, proceed to Figure 2/Scheme 2.
- the log K_{OW} < 3: The bioaccumulation potential in aquatic organisms is considered low, and no further testing is required under any European legislation, except if log K_{OW} ≥ 2 and log K_{OA} is ≥ 5.
- the log K_{OW} ≥ 3-4.5: depending on the relevant regulatory regime and the purpose of the assessment: Start of bioaccumulation assessment.
- the log K_{OW} 3-6: Before proceeding to testing, a bioconcentration factor (BCF) should be predicted using QSAR models (tier 2).
- the log K_{OW} above a value 6: Depending on substance characteristics, experimental bioaccumulation testing may be difficult and produce artefacts. Thus, if experimental work cannot produce valid/acceptable results, other forms of assessments may be necessary for highly hydrophobic (superhydrophobic) compounds. Based on mechanistic models (Larisch and Goss 2017, 2018; Ebert et al. 2023a), it is assumed that superhydrophobic substances (here the assumption is a log K_{OW} > 6) are to be assessed as bioaccumulative as long as no significant biotransformation takes place. This is further explored elsewhere, e.g. in an ongoing UBA project (FKZ 3723 64 4050 “Alternative strategies for the assessment of bioaccumulation of superhydrophobic substances”).

Figure 1: Scheme 1: Proposal of an assessment strategy for bioaccumulation in water-breathing organisms (Mode of aquatic bioaccumulation: Predominantly in lipid).



= Evaluate against regulatory threshold of concern, if specified. Move to higher tier if data is not sufficiently conclusive.

Green bubbles display additional testing options that may not be specifically defined in the respective OECD TGs.

*QSARs for other endpoints (e.g. biotransformation) may also be applicable at this tier in the future
"SP"= Secondary Poisoning
"?" = Inconclusive data

(1) Accumulation not predominantly in lipid. See chapter 4.1.2
(2) Accumulation of ionisable substances. See chapter 4.1.2.1
(3) Accumulation of superhydrophobic substances. See chapter 4.1.1
(4) Evaluation of test data/ selection of test depends on regulatory questions. See chapter 7.2
(5) In case of no aquatic bioaccumulation, move to assessment for air breathing organisms (*cf.* Figure 2).
(6) The "best" BCF is used to evaluate the hazard for Secondary Poisoning in a separate hazard assessment.

Source: own illustration, Fraunhofer IME

Tier 2: In recent years, QSAR models have increasingly been developed for bioaccumulation assessment. If no data is available for the chemical in question, these QSARs should be used – with due regard to the respective domain of applicability – to gain a preliminary understanding of the bioaccumulation potential. If necessary, multiple values from different independent methods (measured and/or estimated) should be obtained to estimate a consolidated BCF (see chapter 5.3).

If the modelled / QSAR-estimated BCF is sufficiently far below any value of regulatory concern, and predictions are shown to be valid and within the applicability domain of respective model further assessment might not be required. If the BCF is at or above the lowest value of regulatory concern, the analysis moves to tier 3. However, because QSAR BCF values have high uncertainty and variability (see chapter 5.3), a margin of safety to any value of regulatory concern should be respected, and there should be sufficient confidence in the QSAR estimates to avoid overlooking potentially bioaccumulative substances despite these uncertainties.

Tier 3 includes OECD TG 319A/B (in vitro substrate depletion assays) or OECD TG 321 (HYBIT) evaluating the bioaccumulation potential using NAMs. If the result of the in vitro test or HYBIT is sufficiently conclusive for the regulatory purpose (risk assessment or PBT assessment) no further test is necessary. If there is insufficient confidence in the result, the other NAMs test should be performed for confirmation. Only if both tests lead to a contradictory result proceed to tier 4 for possible confirmation based on in vivo vertebrate testing (OECD TG 305). ECHA has defined a threshold of concern (see chapter 2.1) for HYBIT. For the in vitro tests thresholds have not yet been established.

The tests in tiers 3 and 4 provide additional testing options that may not be specifically defined in the respective OECD TGs.

In case of a non-B conclusion, it should be further investigated whether there is still a potential for bioaccumulation under terrestrial conditions. If log $K_{OW} \geq 2$, log $K_{OA} \geq 5$, additional assessment is suggested for terrestrial bioaccumulation potential (Scheme 2, Figure 2).

In most cases, Scheme 1 should lead to a conclusive identification of the bioaccumulation potential of a compound that fits the applicability domain. However, always consider available relevant and reliable other data in a WoE approach, such as environmental monitoring data (tier 5) or data from other species and alternative methods (NAMs), which would be outside the scope of this project.

4.1.2 Bioaccumulation not predominantly in lipid

4.1.2.1 Example 1: Ionisable compounds

Tier 1: By definition, the octanol–water partition coefficient (K_{OW}) assumes that the solute exhibits the same speciation in both the water and octanol phases—typically its neutral form. However, processes such as dissociation, can lead to more complex distribution behaviour (Schüürmann et al., 2007; Martin, 2009; Ulrich et al., 2021; Köhler et al., 2023). When multiple species of a solute coexist, their relative proportions collectively determine the apparent partitioning, expressed as the distribution coefficient (log D). For instance, the pH-dependent log D values of acids and bases reflect the balance between their ionised and non-ionised forms, as governed by their p*K*a values and described by the Henderson–Hasselbalch equation.

Tier 2: For ionisable chemicals, the modelled QSAR BCF should be adjusted considering the ionised fraction at the test pH (pH 4–9) using Henderson Hasselbalch and log D. However, it should be noted that p*K*a estimates maybe highly variable (see chapter 5.6).

Tier 3: OECD TG 319A/B: In vitro substrate depletion assays mimic a liver and thus, the pH value of the depletion assays must remain at a physiological state. A change of environmental pH must be introduced by the IVIVE model. In general: Care must be taken that an appropriate IVIVE model that respects the nature of ionisable substances is selected for the extrapolation. The BIONIC model (Armitage et al. 2013) is an example. At the current state, it is only available as Excel sheet for the extrapolation of substrate depletion data derived from S9 assays with rainbow trout.
OECD TG 321: Ionisable substances released into aquatic environments undergo varying degrees of ionisation depending on ambient pH, which can significantly influence their bioaccumulation potential. Bioaccumulation tests, such as OECD TG 321, typically conducted at a single pH value, may therefore underestimate or overestimate bioaccumulation under natural conditions. To address this limitation, exposure conditions need to be established that enable pH stabilisation for the pH-dependent evaluation of bioaccumulation.

Tier 4: OECD TG 305: Ionisable compounds can be tested in flow-through fish tests according to OECD TG 305 with careful experimental modification (e.g., pH control, supplementary log D data). The standard protocol is not fully suitable for accurately assessing the bioaccumulation potential of ionisable compounds without such adaptations.

4.1.2.2 Example 2: PFAS

Perfluoroalkyl substances (PFAS) are regarded as proteinophilic compounds that are rather distributed in biofluids and tissues containing protein and phospholipids. The aquatic bioaccumulation potential should be principally evaluated as described in Figure 1 but based on specific physicochemical properties of PFAS (tier 1) and a novel mechanistic bioaccumulation model for PFAS (Kelly et al. 2024) in tier 2.

Key physicochemical properties (tier 1) are different distribution coefficients that represent the partitioning of PFAS into different biological constituents. Instead of log K_{OW}, the biota-water distribution coefficient (D_{BW}) serves as a key parameter in assessing the likelihood of bioaccumulation of PFAS, integrating both biological and chemical interactions within the ecosystem. D_{BW} is derived through a mass balance approach (Kelly et al. 2024) that incorporates various distribution coefficients representing the equilibrium partitioning of contaminants between biota and water. Specifically, it considers the affinity of PFAS for different biological components and how these relationships influence their concentration in organisms compared to the surrounding water. Several specific distribution coefficients are used, such as: Neutral lipids ($D_{NL\text{-}W}$), Phospholipids (D_{MW}), Albumin ($D_{ALB\text{-}W}$), Transporter proteins ($D_{TP\text{-}W}$), Structural proteins ($D_{SP\text{-}W}$), and Carbohydrates (D_{CW}). The calculation of D_{BW} involves summing the contributions of these various components to estimate the overall distribution of PFAS in the organism. The calculation also includes consideration of the volume fractions of the various constituents in biota, which influence the overall sorptive capacity of the organism. Environmental factors, such as pH, can affect the distribution coefficients, especially for ionisable compounds. Thus, adjustments based on the physicochemical properties of PFAS at the relevant environmental conditions are crucial when estimating D_{BW}. Based on the estimated D_{BW}, whole-body fish BCF values for PFAS can be predicted using the novel mechanistic model (Kelly et al. 2024), Alternatively, fish BCFs for PFAS can be predicted based on a QSPR model (tier 2) recently presented by Kowalska et al. (2024). The QSPR model for PFAS was developed by analysing a dataset of experimental BCF values obtained from laboratory studies. The QSPR model relies on calculated molecular descriptors that quantitatively represent the chemical and physical properties of PFAS molecules. These descriptors include 1D, 2D, and 3D features, such as hydrophobicity, molecular size, and functional group characteristics. Multiple linear regression (MLR) or other statistical techniques are used to correlate the molecular descriptors

with the observed BCF values. This statistical relationship allows the prediction of BCF for new, untested PFAS compounds. The resulting QSPR model can be applied to a large database of PFAS (such as the NORMAN database), enabling the classification of numerous compounds as "not bioaccumulative", "bioaccumulative", or "very bioaccumulative" based on their predicted log BCF values.

If the modelled QSAR-in silico BCF is below the value of regulatory concern, further assessment might not be required. If it is above the threshold, the analysis moves to tier 3 (and 4) as described in Figure 1.

4.1.2.3 Example 3: Surfactants

The aquatic bioaccumulation potential should be principally evaluated as described in Figure 1 but based on specific physicochemical properties of surfactants (tier 1) and a novel model for BCF prediction (McLachlan et al. 2023) in tier 2. Due to the specific phase behaviour of surfactants and their full ionisation under test conditions, these substances are outside the applicability domain of most tools used to calculate log K_{OW}. Alternative partition coefficients that may be more biologically relevant such as $K_{membrane(lipid)/water}$ or $K_{liposome/water}$ have been measured and modelled (Hodges et al. 2019, Potter et al. 2021, Qin et al. 2023, Droge et al. 2021). The bioaccumulation assessment of surfactants principally follows the previously described assessment schemes with a tiered assessment process relying on a combination of physicochemical data, model-based prediction of BCFs, NAMs, and in vivo testing. Thresholds for BCF assessment agree with the assessment concept presented in Scheme 1 (Figure 1). However, initial evaluation is based on specific physicochemical properties (tier 1). Here, fundamental properties like p*K*a (acid dissociation constant), partitioning (e.g., log D), and the membrane lipid/water distribution ratio (D_{MLW}) are examined. If D_{MLW} is below a certain threshold (to be discussed elsewhere), no further testing is required. Otherwise, a BCF needs to be predicted in tier 2 according to McLachlan et al. (2023). If substances show a modelled BCF above the regulatory threshold value, further testing in tier 3 (and 4) is required as described in Scheme 1 (Figure 1).

4.2 Integrated assessment of bioaccumulation in air-breathing vertebrates

Chemicals that are sufficiently volatile will be readily eliminated by exhalation by air-breathing organisms, as expressed by the octanol-air partition coefficient (log K_{OA}) of the compound (see chapter 5.5). If a substance is not bioaccumulative for aquatic organisms (gill breathers), still an assessment of the biomagnification potential in vertebrates of terrestrial food chains and air-breathing marine wildlife as well as in humans needs to be conducted, as certain substances do not bioaccumulate in aquatic food-webs as much as in air-breathing animals (e.g., terrestrial mammals, birds).

A concept for bioaccumulation assessment in air-breathing mammals was presented by Lee et al. (2022), Wania et al. (2022) and ECHA (2022). Screening of chemical substances for biomagnification in air-breathing organisms based on K_{OA} and K_{OW} is easy to be achieved for those substances where these are relevant, as described in the previous chapters. However, this approach can be expected to generate many false positives. Because many of these high-K_{OA} and high K_{OW} chemicals can be biotransformed in organisms, their biomagnification potential is reduced (Lee et al. 2022). To decrease the number of false positives it is also important to account for dissociation and the formation of charged species which can affect the elimination of organic chemicals from air-breathing organism in two ways: I) increase of water solubility and II) increase of volatility (Wania et al. 2022). In silico predictions of biotransformation loss can help to identify substances with low, medium and high biomagnification potential which should be further tested in an in vitro rat liver S9 or hepatocyte biotransformation assay. Following the

determination of in vitro biotransformation constants, whole-organism biotransformation rate constants and half-lives or BMFs in rats can be calculated (Lee et al. 2022). A BMF ≥ 1 leads to a B or vB classification, but no decision has been taken whether the BMF, the elimination half life or the biotransformation rate will be used as a regulatory criterion. Also, the relevant models need further validation. The in vitro rat liver S9 or hepatocyte biotransformation assay has thus far been applied mostly to neutral hydrophobic organic substances and pharmaceuticals. Therefore, the potential of the assay for the testing of difficult to test substance classes (e.g. PFAS, siloxanes) needs to be investigated.

4.2.1 Bioaccumulation predominantly in lipid

A process for assessing bioaccumulation of neutral organic chemicals, focusing on bioaccumulation in air-breathing organisms was proposed by ECHA (2022). The bioaccumulation assessment of PFAS and surfactants was not within the scope of this project.

To proceed with further testing following aquatic bioaccumulation assessment (Scheme 1), a tiered approach for assessing chemical bioaccumulation in air-breathing organisms proposed by ECHA (2022) should follow as outlined in "Scheme 2" (Figure 2).

The proposal for the tiered approach is summarised as follows:

Tier 1: Screening assessment:

- log K_{OW} > 2: This suggests moderate to high hydrophobicity and
- log K_{OA} > 5: Indicates the chemical's potential for partitioning into organisms.

Tier 2: In silico biotransformation:

- Estimated biohalf-life ($HL_{biotransformation}$) > 50 days in humans or 4 days in rats (ECHA 2022) using in silico half-life estimates. Other taxa could be used if sufficient data is available and regulatory acceptance can be found. For the models, regulatory consensus also has to be found.

Tier 3: Refined screening assessment using alternative testing methods

- Use rat liver S9 or hepatocyte in vitro substrate depletion assays that are specifically designed for bioaccumulation screening of neutral hydrophobic organic chemicals with bioaccumulation potential in air-breathing organisms. S9 assays are described by Lee et al. (2017, 2022). An in vitro depletion rate constant threshold value of 10 x 0.3 hrs^{-1} or 3 hrs^{-1} (corresponding to an in vitro half-life time of 0.23 hrs) is suggested as a possible conservative screening value for rat liver in vitro S9 depletion tests (ECHA 2022). Depletion rates could also be extrapolated to biohalf-lives which could then be compared against the thresholds presented in tier 2. There is currently limited experience with either approach, therefore further evaluation is needed.
- Extrapolation of rat liver S9 in vitro bioassay data to a BMF based on the methods developed by Lee et al. (2017, 2022), or extrapolation of rat hepatocyte bioassay data to a BMF using an appropriate IVIVE model such as the B-compass rat model (Krause & Goss 2021). There is currently limited experience in this regard, therefore further evaluation is needed.

Tier 4: Refined assessment using in vivo testing methods.

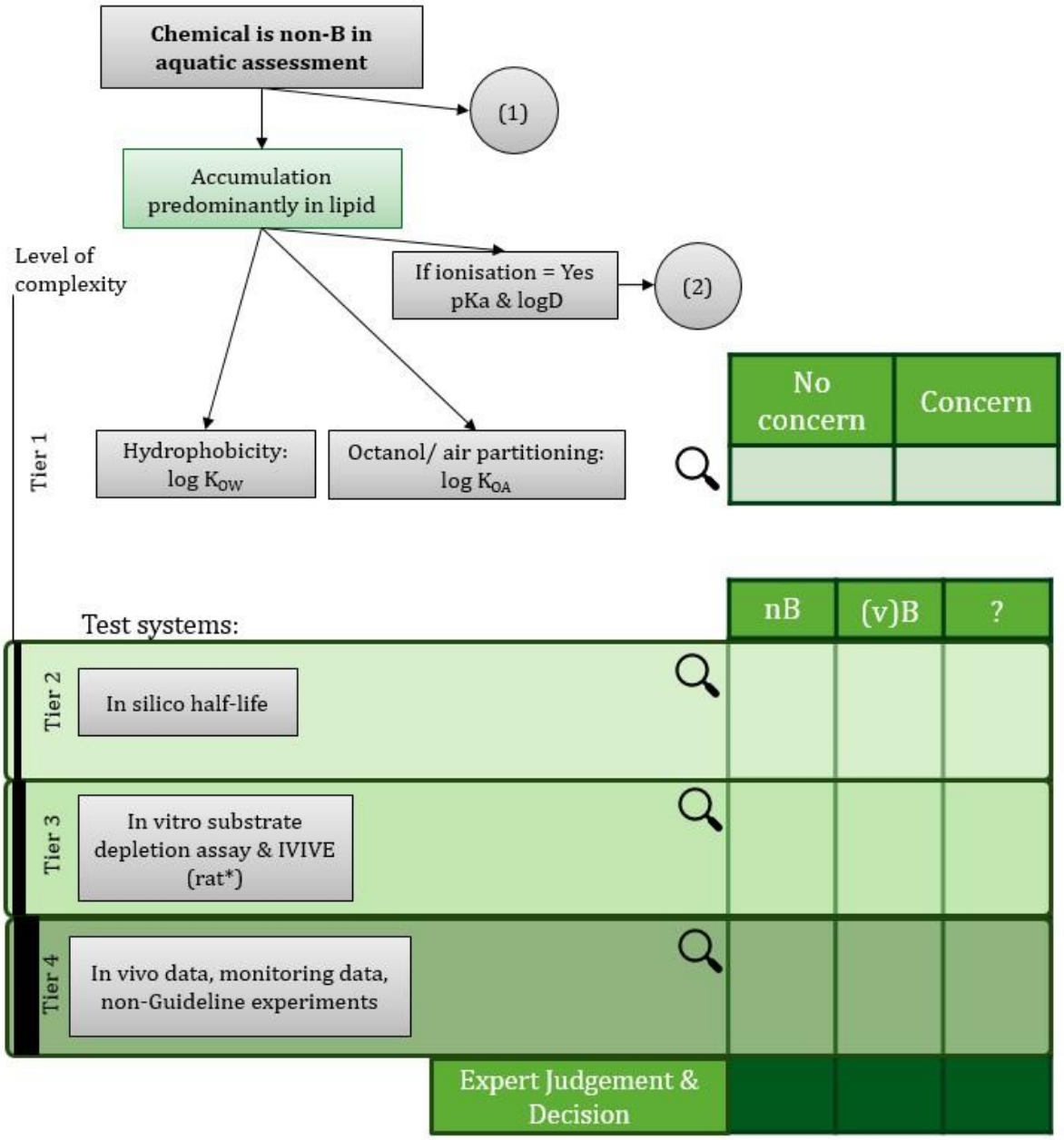
- In vivo testing is required when no definitive conclusion on bioaccumulation can be drawn from the previous tiers. Preferably, available toxicokinetic data should be used. If available,

reliable quality in vivo field data, in particular high-quality trophic magnification factors (TMF) may also be considered.

There are calculations for S9 data that show, that if the rate constant for depletion is equal to or above 3 h^{-1}, the chemical is not expected to biomagnify in rats (ECHA 2022), and further testing may not be necessary. This rate is not applicable to depletion rates determined using hepatocytes. Alternatively, there may be an option to extrapolate depletion data of S9 and/ or hepatocytes to biohalf-lives using an appropriate IVIVE model in conjunction with mass-balance models, mass-transfer models, or ideally, single or multi-compartmental physiologically based kinetic (PBK) models. These biohalf-lives could be compared against the thresholds in tier 2 (which require validation). If the in vitro assay suggests potential for bioaccumulation, an IVIVE can also be performed to extrapolate in vitro data to biomagnification factors (BMFs). In this case a BMF > 1 indicates bioaccumulation, and the chemical is assessed as B (bioaccumulative), or even vB. At the current state, there is not yet sufficient experience to deduce classifications. If the results of the test are inconclusive, or if it is not possible to conduct an in vitro test due to the characteristics of the test substance, other information on the bioaccumulation potential of the substance should be investigated and evaluated in a WoE approach (tier 4).

Note: The refined assessment of bioaccumulation in air-breathing organisms is in a proposal state. It has not been validated and a validation is not within the scope of this project. More information on the current state of assessing bioaccumulation in air-breathing organisms can be found in the ECHA (2022) report.

Figure 2: Scheme 2: Proposal of a testing and assessment strategy for bioaccumulation in air-breathing organisms.



= Check against regulatory threshold of concern, if specified. Move to higher tier if required.

*depending on protection goal and technical developments, other species can be used instead/ additionally.

"?" = Inconclusive data

(1) Accumulation not predominantly in lipid. See chapter 4.1.2
(2) Accumulation of ionisable substances. See chapter 4.1.2.1

Source: own illustration, Fraunhofer IME

5 Evaluation of test systems

This chapter presents the results of a comprehensive method analysis, offering a structured comparison of traditional (OECD TG 305) and alternative tests, namely *Hyalella azteca* tests (OECD TG 321) and in vitro substrate depletion assays (OECD TG 319A/B; in vitro rat liver S9 or hepatocyte assay), as well as BCF-QSAR models and physicochemical screening parameters for bioaccumulation assessment. Methodological uncertainties are identified, and opportunities for improving bioaccumulation assessments within a contemporary regulatory context are outlined. The results of this work on log K_{OW} as a physicochemical parameter for assessing bioaccumulation were published in Nendza et al., 2025.

5.1 Data collection

The data collection was conducted to provide a sufficient number of data sets for the case studies. The case studies were carried out to compare the results of NAM studies with results of in vivo fish bioconcentration studies. Therefore, the literature was reviewed to identify substances with available information on bioaccumulation potential obtained from NAM-based tests, including both in vivo methods (e.g., *Hyalella azteca* BCF studies) and in vitro approaches (e.g., depletion assays using hepatic S9 fractions or hepatocytes), in combination with IVIVE methodologies (e.g., Nichols et al. 2018a, Lee et al. 2022). Substrate depletion data for more than 200 substances which were derived using hepatic trout material (hepatocytes and S9) were compiled. Additional data from hepatic S9 and hepatocyte tests with rainbow trout were obtained from the EURL ECVAM Fish In Vitro Intrinsic Clearance Database (Halder et al. 2018). A search for substrate depletion data generated with hepatic material (S9 and/ or hepatocytes) from rats was also conducted and delivered data for approx. 40 substances. In vivo fish BCF data was curated from different sources: Preceding literature searches (Schlechtriem et al. 2019), accessible training data for QSAR models (EpiSuite data, VEGA data, EURAS data) and directed searches for BCF values of substances for which alternative laboratory bioaccumulation endpoints are available. Furthermore, the OECD QSAR Toolbox was used to search for BCF endpoint data of specific compounds. If available, BCF data from ECHA Dossiers was also used.

A sufficiently large dataset was required to evaluate methods for the estimation of the physicochemical parameters log K_{OW}, log K_{OA} and p*K*a, but also for the calculation of QSAR BCFs. Based on the literature search, 231 chemicals were selected, representing diverse chemical classes, such as POPs, PCB, PAH, flame retardants, pesticides (insecticides, herbicides, fungicides, …), pharmaceuticals (analgesics, antibiotics, antidepressants, …), fragrances, biocides, UV-filters, plasticisers. For some pharmaceuticals, different enantiomers were considered. Also included were examples of substances such as PFAS, siloxanes and surfactants with both the neutral and ionic forms of anionic surfactants (Nendza et al. 2025). Based on this list of substances, with a broad coverage of applicability domains and influential physicochemical properties, a systematic consideration of the log K_{OW} values of the identified substances was carried out with the help of various, established QSAR models. A consensus modelling for consolidated log K_{OW} values was developed. In addition, the use of the physicochemical parameters log K_{OA} and p*K*a was evaluated. In addition, estimation methods for the derivation of log BCF-fish QSAR data were described and their uncertainties as well as possibilities of consolidation were presented.

For the comparison of endpoints, also additional laboratory tests were conducted to complete some of the data sets where experimental NAMs data was missing. Conducting the studies also allowed the test systems to be compared in terms of their advantages, disadvantages and limitations for determining the bioaccumulation potential of substances. For the case studies, a set of 26 substances (Table 3) was selected, based on data availability, to test methodological boundaries.

Table 3: Substances for evaluation of the test strategies.[1)]

Name	Molecular weight (g/mol)	log K_{ow}	log K_{OA}	p*K*a (acid)	p*K*a (base)
Simazine	201.66	2.12	8.94	--	1.7
Azoxystrobin	403.39	3.00	13.9	--	0.6
GenX	347.08	3.26	5.09	-0.03	--
Terbutryn	241.36	3.41	9.83	--	-1.5
17α-Ethinylestradiol	296.41	3.84	13.6	10	--
β-Hexachlorocyclohexane	290.81	3.87	7.59	--	--
Prochloraz	376.66	3.91	12.2	--	5.7
Fluorene	166.22	4.06	6.69	--	--
Fluoxetine	309.33	4.33	10.0	--	9.6
Diclofenac	296.15	4.35	12.9	4.4	--
Anthracene	178.23	4.41	7.44	--	--
PFOS	500.13	4.42	5.81	-5.7	--
Phenanthrene	178.23	4.43	7.45	--	--
Pentachlorophenol	266.32	4.81	9.31	4.5	--
Lauric acid	200.32	4.84	8.51	4.9	--
Chlorpyrifos	350.57	4.89	9.47	--	--
Pyrene	202.26	4.91	8.75	--	--
Triclosan	289.54	4.91	11.08	8.8	--
Octamethylcyclotetrasiloxane (D4)	296.62	4.98	4.13	--	--
Trifloxystrobin	408.38	5.08	11.2	--	--
Pentachlorobenzene	250.32	5.11	6.53	--	--
Methoxychlor	345.64	5.25	10.6	--	--
o-Terphenyl	230.31	5.58	8.80	--	--
Benzo[a]pyrene	252.32	6.07	11.0	--	--
PCB 153	360.86	7.08	10.0	--	--
UV-234	447.58	7.68	15.32	10.3	--

[1)] Consolidated values of log K_{OW} and log K_{OA}. The p*K*a values predicted by the ACD/Labs Galas tool are included in the table. For details see chapters 5.4 - 5.6.

5.2 New Approach Methodologies (NAMs) for bioaccumulation assessment

In recent years, a number of NAMs were developed as alternatives to the fish BCF test (OECD TG 305 I). There is a strong commitment from all stakeholders for the implementation of NAMs in chemical risk assessment to ultimately replace animal testing as expressed during ECHA's New Approach Methodologies Workshop "Towards an Animal Free Regulatory System for Industrial Chemicals" 31 May – 1 June 2023, Helsinki, Finland (ECHA 2023d).

5.2.1 OECD TG 321: *Hyalella azteca* Bioconcentration test (HYBIT)

The *Hyalella azteca* Bioconcentration Test (HYBIT) which is carried out according to OECD TG 321 (OECD 2024a) provides a non-vertebrate test for bioconcentration in aquatic environments. It provides an established endpoint (BCF) and at the same time advantages for laboratory testing compared to OECD TG 305 such as simple cultivation, short generations, shorter time to steady-state, compact test system. The OECD TG 321 (OECD 2024a) studies can be performed under flow-through or semi-static conditions which should allow stable test item concentrations over time. Additional information regarding literature background for the HYBIT are provided in chapter A.1.1 of the Annex.

5.2.1.1 Experimental evaluation of the test system

Table 4Table 4 summarises the substance selection. The material and methods of the HYBIT experiments conducted as part of this project are provided in the Annex in chapters A.1.1 – A.1.4. Details of preliminary experiments are given in chapter A.1.5. Results are presented in chapters A.2.1 – A.2.8 of the Annex.

Table 4: Substances tested in HYBIT experiments as part of this project.

Substance	Substance characteristics	Available data	Reason for testing
Pyrene	Neutral organic 4.5 < log K_{OW} <6	*H. azteca* BCF (^{14}C TRR-based) 319A & 319B Rat S9 depletion data Fish in vivo BCFs	Reference substance
β-HCH	Neutral organic 3 < log K_{OW} < 4.5	Rat S9 depletion data Fish in vivo BCFs	Proof of principle: NAMs lead to the same conclusions as in vivo data (not B in water-respiring organisms, B in air-breathers)
UV-234	Neutral organic Log K_{OW} > 6	*H. azteca* BCF	Reproducibility Test limits
D4	Neutral organic 4.5 < log K_{OW} < 6 Volatile	Fish in vivo BCFs 319B (published after project work)	Test limits

Substance	Substance characteristics	Available data	Reason for testing
Diclofenac	Ionisable organic	*H. azteca* BCF 319A Rat HEP depletion data Fish in vivo BCFs	Reproducibility Testing at different pH values Test limits
Fluoxetine	Ionisable organic	Fish in vivo BCFs (at different pH values)	Testing at different pH values Test limits

Neutral organic chemicals: In this study flow-through tests were carried out with a moderately (β-hexachlorocyclohexane) and a highly hydrophobic (pyrene) neutral organic chemical. Both tests were concluded successfully (see Annex A.2 for further details).

Chemicals with elevated hydrophobicity ("superhydrophobics"): In the study with UV-234, steady state was not conclusively reached after 10 days of exposure, as suggested by tissue concentrations measured toward the end of the uptake phase. Consequently, the resulting steady-state BCF (BCF_{SS}) values must be interpreted with caution, underscoring the need for further refinement of test strategies for superhydrophobic substances. This topic is further explored in an ongoing UBA project. For details on the UV-234 experiment conducted in this project see Annex A 2.3.

Schlechtriem et al. (2022) tested the solvent-facilitated application of UV-234 in a HYBIT flow-through BCF test resulting in a lipid-normalised kinetic BCF of 902 (3%, w/w). Since biofilm formation was observed, it could be only speculated how this compound would behave in a solvent-free bioconcentration test with *H. azteca*. In this study, the BCF test with UV-234 was repeated. Different exposure methods including column generated concentrations and supernatant solutions (stirring bottle) were prepared. However, stable exposure conditions could only be reached by using solvents. Different solvent systems were tested with DMSO at 0.002% leading to the best result. In this study, a lipid-normalised kinetic BCF of 2063 (3%, w/w) was determined which is significantly higher compared to the former result which, when normalised to the same lipid content of 3% (w/w), amounts to a BCF_{kL} of 901. With an uptake rate constant k_1 of 1295 L/(kg*d), UV-234 showed a much faster uptake compared to the first study where a k_1 of 198 L/(kg*d) was determined. The elimination rate constants of both studies were 0.273 [$1d^{-1}$] and 0.9288 [1^{-1}] for the first and second study, respectively. Ebert et al. (2023a) developed mechanistic model for predicting the uptake and elimination rates of chemicals in *Hyalella azteca*, specifically focusing on bioconcentration factors (BCFs). The model developed for *Hyalella azteca* helps to expand the understanding of the bioconcentration mechanisms of chemicals in aquatic invertebrates. While both *H. azteca* and fish tests provide valuable insights into bioconcentration of superhydrophobic compounds, it is crucial to consider the specific characteristics of the test systems and the physiological characteristics of the organisms to interpret the results correctly.

In a further BCF test, D4 (octamethylcyclotetrasiloxane), a superhydrophobic and volatile substance that belongs to the class of cyclic volatile methyl siloxanes (cVMS) was tested. The substance is known to hydrolyse in water (Brooke et al. 2009). Biofilm formation was observed following the use of acetone as solvent. An application of D4 in a stable manner for a HYBIT BCF test was not possible (see Annex A.2 for further details). Further experiments would have been

necessary to develop techniques that permit the establishment of stable exposure conditions, which were beyond the scope of this project.

Experimental evaluation (ionisable neutral organics): In this project two HYBIT BCF tests were performed at varying pH values. The HYBIT BCF studies with diclofenac and fluoxetine, in which both substances were tested a 2 pH levels each, showed that the test system is suitable for testing ionisable organic compounds, allowing to investigate the complex interactions between pH, ionisation, and bioaccumulation potential of such compounds. Experiments that were conducted in this project demonstrated that a HYBIT test can principally be performed at varying pH values, both technically and physiologically. The technical part requires fine tuning to fully eliminate the singular malfunction events that occurred in two of the four experiments. It was possible to perform tests in a pH range from 5 – 9. The hypothesis that the pH of the surrounding medium has an impact on the bioaccumulation potential was confirmed (see Annex A.2.4 to A.2.7 for further details).

Fluoxetine is biotransformed to its pharmaceutically active metabolite norfluoxetine, which displays an increased hydrophobicity (Oakes et al. 2010). A fish bioconcentration study with fluoxetine showed that norfluoxetine concentrations in fish are as high or even higher than fluoxetine concentrations (Nakamura et al. 2008). At lower pH values, the norfluoxetine to fluoxetine ratios were higher than at a pH value of approx. 9. The overall concentration of fluoxetine and norfluoxetine in the fish tissues increased with increasing pH of the test medium. The same observations can be made in the *H. azteca* data of this project (*cf*. Supplementary Information "HYBIT raw data").

Therefore, the HYBIT is a tool that can be used to evaluate the bioaccumulation potential of ionisable substances in aquatic organisms. Additional experiments will be necessary to exclude that the buffer substances or the changed pH values of the exposure medium impact the bioconcentration potential of the tested substance. One way to do this could be to perform test with a benchmarking substance at different pH levels and using different buffer substances. The resulting BCF values should then be compared.

5.2.1.2 Conclusions from experimental and literature data

Laboratory studies conducted in accordance with OECD TG 321 as part of this project have confirmed the suitability of the test system for flow-through bioconcentration tests. Also testing at different pH values is possible. However, in this case the validity criteria of the test guideline were not met, as isolated incidents led to fluctuations in water concentrations and, at one point, to mortality. The technical implementation of pH variation may require further optimisation, but the proof of principle for testing under different pH conditions was successfully demonstrated.

Testing of superhydrophobic substances is possible as well, though the complexity of application strongly depends on the specific compound. For example, UV-329 could be applied consistently using a column elution method (Schlechtriem et al. 2022), whereas UV-234 could not be applied steadily with the same approach. When interpreting the resulting HYBIT BCF values for highly hydrophobic substances, possible effects arising from experimental challenges should be taken into account. Testing such compounds, solvents should be used with caution. Use of solvents can promote biofilm formation, and the potential impact of biofilm on bioconcentration in *H. azteca* has not yet been investigated.

5.2.2 OECD TG 319A/B: In vitro substrate depletion assays

In vitro substrate depletion assays that are carried out according to OECD TGs 319A/B provide an in vitro rate for the biotransformation of the test chemical. This rate can be used to refine in silico predictions of the bioconcentration potential, which is referred to as IVIVE (In Vitro to In

Vivo Extrapolation). Advantages of these tests are fast experiments, high-throughput testing options (for tests with S9 fractions) and only small quantities of the test substance are required. More details are provided in chapter A.3 of the Annex.

5.2.2.1 Experimental evaluation of the test system

Substrate depletion assays with trout hepatocytes were performed for the substances pyrene, β-HCH, UV-234, and D4. Table 5 summarises the substance selection. The chapter A.4.2 in the Annex can be consulted for further details on the test designs. BCF values were estimated by applying the B-compass fish IVIVE model (details in Annex A.4.4). Details for the results are reported in chapters A.6 (substrate depletion assays) and A.8 (IVIVEs).

Pyrene was readily biotransformed by the rainbow trout hepatocytes. No biotransformation was detectable for the test substances β-HCH and UV-234 under the applied test conditions. It was not possible to test D4. Cantu et al. were able to test D4 using rainbow trout S9 fractions (Cantu et al. 2025). It is possible, that the headspace in the test vessels that were used in this project was too large during the preliminary tests in this project.

Table 5: Substances tested in OECD TG 319A assays as part of this project.

Substance	Substance characteristics	Available data	Reason for testing
Pyrene	Neutral organic 4.5 < log K_{OW} < 6	*H. azteca* BCF (^{14}C TRR-based) 319A & 319B Rat S9 depletion data Fish in vivo BCFs	Reference substance
β-HCH	Neutral organic 3 < log K_{OW} < 4.5	Rat S9 depletion data Fish in vivo BCFs	Proof of principle: NAM lead to the same conclusions as in vivo data (not B in water-respiring organisms, B in air-breathers)
UV-234	Neutral organic Log K_{OW} > 6	*H. azteca* BCF Fish in vivo BCF	Test limits
D4	Neutral organic 4.5 < log K_{OW} < 6 Volatile	Fish in vivo BCFs 319B (published after project work)	Test limits

5.2.2.2 Conclusions from experimental and literature data

The determination of an IVIVE BCF for the assessment of the bioaccumulation potential of substances comprises two parts: The substrate depletion assay and the IVIVE extrapolation. For every test chemical, it is necessary to assess the applicability domain independently for both the depletion assay and the IVIVE model. Published data shows that approx. 270 substances could be tested using rainbow trout liver S9 fractions and/ or hepatocytes. For two substances it was explicitly reported that no first order depletion could be determined (prochloraz in Kosfeld et al. 2020 and o-terphenyl in Fay et al. 2014b). In those cases where no depletion could be

determined it is usually not known if the substance cannot be biotransformed or if the assay conditions were not ideal for the test substance. If no biotransformation can be determined in a substrate depletion assay, the IVIVE BCF is solely based on the IVIVE model, its parameterisation and the substance-specific the input data (e.g. log K_{OW}).

IVIVE models (input: log K_{OW})

In silico extrapolations are commonly performed using log K_{OW} based extrapolation models. This project has shown that measured and calculated log K_{OW} values reveal considerable variability up to one log unit and more, without systematic pattern (see chapter 5.4). Consensus modelling for consolidated log K_{OW} values is recommended to derive more robust and reproducible values (for details, *cf*. chapter 5.4), since the log K_{OW} value can have a significant impact on the results of log K_{OW} based IVIVE models. Looking at BCF values reported in the literature, a large range of in vitro BCF values can be seen for single compounds, e.g. benzo[a]pyrene (BCF = 930-5923) or methoxychlor (BCF = 423-4503). Apart from biological variability, one factor could be the assay condition, the other factor could be the model selection and parameterisation (*cf*. SI "Substance data endpoints"). Standardisation of the extrapolation procedures and perhaps consolidation of the in vitro BCF values may help to get more robust and reproducible results, as the determination of depletion rates is described in detail in OECD TGs 319A/B. Current work aims at harmonising the IVIVE models for OECD TG 319 A/B, which involves the revision of the Guideline's Guidance Document (OECD 2018c).

Substrate depletion assay: Testing superhydrophobic compounds

The suitability of the in vitro depletion approach for testing superhydrophobic compounds needs to be further investigated. The test chemicals once used to evaluate the reliability of the in vitro depletion assay (ring trial) were characterised by a moderate to high hydrophobicity. Only one substance (deltamethrin) was tested with a log K_{OW} > 6. (log K_{OW} = 6.2) (Nichols et al. 2018a). This dataset was enlarged over the years, by testing substances with elevated hydrophobicity, such as siloxanes (Cantu et al. 2025), or PCB 153 (Trowell et al. 2018).

Substrate depletion assay: Ionisable substances

Depletion assays determined in this project and found in the literature have also been conducted with ionisable substances, mainly pharmaceuticals. Based on the observation that the state of charge a molecule exhibits also influences its partitioning behaviour into lipids (Köhler et al. 2023), extrapolations for ionisable substances were performed using log K_{OW}-based extrapolation models and the log D of both the neutral and the ionised molecule. This is not state of the art and a model designed for the extrapolation of ionic substances, such as the BIONIC model, should be used instead (Armitage et al. 2013). This model requires additional partitioning factors. Retrieving those factors in a similar manner as it was done for the log K_{OW} was beyond the scope of this project. Furthermore, the available Excel sheet only allowed the input of S9 depletion rates, not hepatocyte depletion rates. Since data from both test systems should be compared using the same model, the BIONIC model could not be used within this project. As can be expected, the IVIVE BCFs are higher if the inserted log D value is also higher. For fluoxetine, in vivo fish BCF_{SS} data (not lipid-normalised, unknown if steady-state was reached) are available and broadly match in vitro results, though they are limited in quality and from medaka, not rainbow trout (Nakamura et al. 2008). For diclofenac, no pH-dependent fish bioconcentration data are available. Generally, where data or models are lacking, log D values for ionised and neutral forms can indicate worst- and best-case bioconcentration. Ionisable substances may behave differently from neutral organics, and such differences, including ion-trapping (*cf*. Köhler et al. 2023), are not captured in log K_{OW}-based models.

IVIVE models for chemicals other than neutral organic compounds

All log K_{OW}-based extrapolation models benefit from a broad dataset of in vivo fish BCF values that were determined over decades. These data were useful for model development and also validation. Whether similar datasets will be necessary to validate extrapolation models for other substance classes (e.g., surfactants, PFAS, ...) is not clear yet. In any case, the predictive power of new models will have to be evaluated. For a broad application, suitable models should be easily available and easy to use to reach a broad acceptance. Regulatory application requires the models to be validated and transparent in their underlying assumptions and applicability domain.

IVIVE for different fish species

OECD TG 305 allows in vivo testing with a set of different fish species. In case this species variability should prospectively be represented by OECD TG 319A/B, the following may need to be considered. IVIVE models have so far been focused on rainbow trout. The metabolism of different fish species may vary depending on the substance (e.g., Roberts et al. 2011, Bischof et al. 2016), accordingly these differences should be respected by the extrapolation models, as well. The adaptation of hepatocyte (Bischof et al. 2022) and S9 (Zercher et al. 2024) protocols to other fish species has been demonstrated. If the OECD TG 319A/B should be expanded to other fish species, there is a need for refined extrapolation models for these other fish species (Zercher et al. 2024, Bischof et al. 2022). In a regression-based BCF extrapolation approach, such a species refinement was applied for carp (*cf*. Laue et al. 2023).

5.2.3 In vitro rat liver S9 and primary hepatocyte biotransformation assays

The assessment of chemicals for their bioaccumulation potential is commonly assessed for water-respiring organisms. These assessments are, however, not necessarily protective for air-breathing organisms. Since no new vertebrate tests should be performed, NAM methods are required. Thus far, the most promising approach is a substrate depletion assay using hepatic material of rats. This in vitro biotransformation rate could be used to refine an in silico rat BMF. Details are provided in chapter A.5.1.

5.2.3.1 Experimental evaluation of the test system

Experimental evaluation: In vitro BMF values were obtained by performing the depletion assay with rat hepatocytes with the neutral, organic substances pyrene, chlorpyrifos, UV-234 as shown in Table 6Table 6. Chapter A.5 of the Annex describes the materials and methods for the tests, chapters A.7 and A.9 describe the experimental and IVIVE results, respectively. Rat hepatocytes were employed to broaden the data basis of the test system and, where applicable, to better understand its limitations. Pyrene and chlorpyrifos were both readily biotransformed by the rat hepatocytes. No biotransformation was detectable for UV-234 under the applied test conditions. It was not possible to test D4.

Table 6: Substances tested in substrate depletion assays with rat hepatocytes as part of this project.

Substance	Substance characteristics	Available data	Reason for testing
Pyrene	Neutral organic 4.5 < log K_{OW} < 6	*H. azteca* BCF (^{14}C TRR-based) 319A & 319B Rat S9 depletion data	Reference substance

Substance	Substance characteristics	Available data	Reason for testing
		Fish in vivo BCFs	
Chlorpyrifos	Neutral organic 4.5 < log K_{OW} < 6	*H. azteca* BCF 319B Fish in vivo BCFs	Proof of principle: Increased accumulation potential in water respiring organisms compared to air-breathers
UV-234	Neutral organic Log K_{OW} > 6	*H. azteca* BCF	Test limits
D4	Neutral organic 4.5 < log K_{OW} < 6 Volatile	Fish in vivo BCFs 319B (published after project work)	Test limits

5.2.3.2 Conclusions from experimental and literature data

The depletion assay with rat hepatocytes was evaluated with neutral organic substances (pyrene, chlorpyrifos, UV-234) during this project. Pyrene was tested in rat S9 fractions by Lee et al. (2022) and provides a direct comparison option. BMF values that were extrapolated from in vitro biotransformation data determined in this project using the B-compass Rat model (Krause & Goss 2021) were higher (0.773) than those determined in Lee et al. (2022) extrapolated with their own model with a value of 0.022. Both values were below a BMF of 1. This allows the cautious conclusion that rat hepatocytes may generally be suitable to derive hepatic biotransformation data in vitro in a similar manner as hepatic S9 fractions of rat. The following aspects may be necessary to increase the reliability and reproducibility of rat hepatocytes as model system:

- The extraction of hepatocytes can require more work compared to the extraction of S9 fractions. A direct protocol transfer of the one that has been described for S9 fractions (Lee et al. 2022) was not possible in this project.
- The extracts of rat hepatocytes caused more matrix effects in LC-MS analyses compared to rainbow trout hepatocyte extracts. The test substances UV-234 and pyrene were both tested using trout and rat hepatocytes. In the case of pyrene (GC-MS analyses), trout hepatocytes could be used to create matrix calibration samples. In the case of UV-234 (LC-MS analyses), rat hepatocytes had to be used for matrix calibration as trout hepatocytes showed significant differences. Also with rat hepatocytes, some of the quality control samples of run 1-3 did not match the calibration probably due to matrix effects. Analyses for chlorpyrifos were only reliable when measured with as much dilution as possible. This indicated that compounds that interfere with the analysis (e.g., salts) were co-extracted from rat hepatocyte samples, possibly from the specific buffers.

Accordingly, refinement steps for in vitro depletion assays with rat hepatocytes could include the following aspects:

- Medium selection and supplementation
- Incubation in a CO_2-supplemented environment

► Adaptions/ refinements of sample extraction and/ or analytical methods

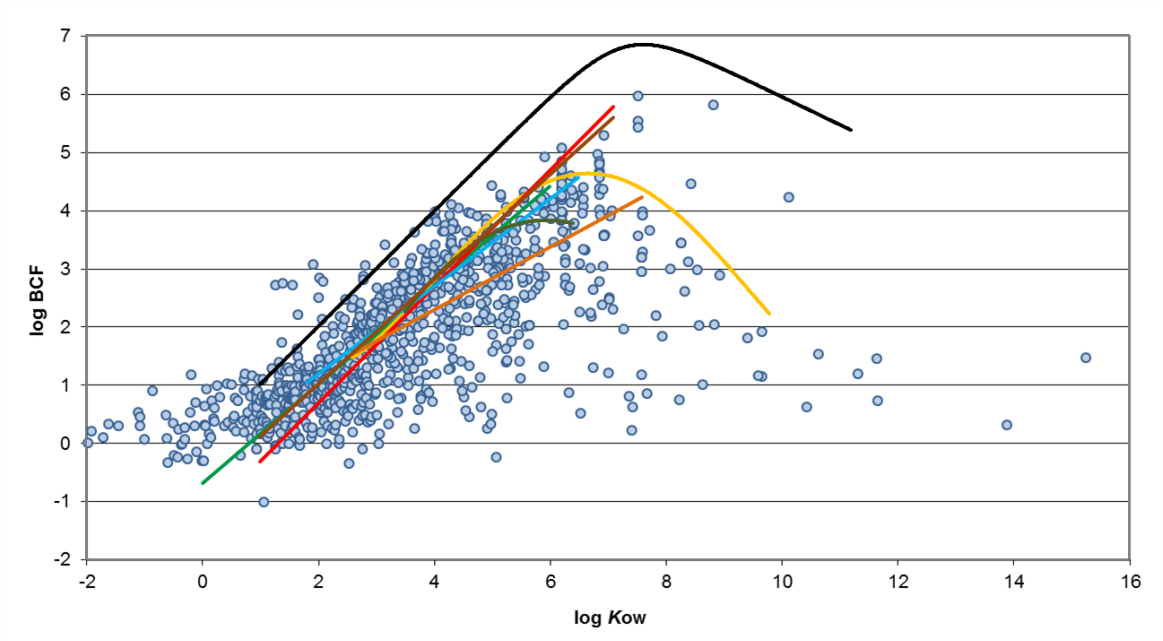
The selection of the incubation medium and its supplementation has an impact on cell viability and the biotransformation capacities of the cells. A sufficient biotransformation of the test compounds can only be obtained under ideal physiological conditions. However, it should also be considered that medium components can also interfere with the downstream analyses of the extracted sample. As such, a balanced medium selection has to be determined. Nonetheless, it is also possible that sample extraction methods may need refinement if vital medium components interfere with the analysis. Cell viability may also be improved by incubating cells in an incubation environment that provides a 5% CO_2 saturation. This is common practice in the cell culture of mammalian cells. Based on the review by Louisse et al. (2020), it was decided that this additional effort may not be necessary, as approx. 65% of publicly available depletion data generated using rat hepatocytes was not obtained under such incubation techniques.

The literature search pointed towards a need for harmonisation of in vitro depletion assays using rat and human hepatocytes (Louisse et al. 2020). Even though only a few new data were contributed, the results from this project support these findings. The development of an OECD test guideline would support the systematic integration of the test into regulatory frameworks.

5.3 Quantitative structure-activity relationships (QSARs)

QSARs aim to describe and predict biological activities of substances from their chemical structures. Over the past 50 years, considering bioconcentration as a distribution between aqueous and organic phases has produced many direct relationships with log K_{OW} (Figure 3).

Figure 3: **Relationships between experimental fish bioconcentration data and log K_{OW}. Examples of BCF-QSAR models: black: (Nendza 1991) bilin worst case, green: (Veith et al. 1979) TGD eq. 74, yellow: (Connell & Hawker 1988), red: (Mackay 1982), blue: (Schüürmann & Klein 1988), olive green: (Könemann & van Leeuwen 1980), orange: (Neely et al. 1974), brown: (Lu et al. 1999).**



Source: own illustration, AL-Luhnstedt

5.3.1 QSAR models for bioaccumulation assessment

Bioaccumulation is governed by four major processes (Lyman et al. 1982; Pfeifer et al. 1984; Hodgeson and Levi 1994; Nendza 1998; Müller and Nendza 2007): Absorption, Distribution, Metabolism and Excretion (ADME). While absorption and distribution increase bioaccumulation, metabolism and excretion reduce bioaccumulation.

Numerous QSARs make use of the fact that bioaccumulation of stable substances is determined by partitioning between aqueous and lipid phases, e.g. (Lyman et al. 1982; Connell and Schüürmann 1988; Nendza 1991; Bintein et al. 1993; Nendza 1998; Dimitrov et al. 2003; Papa et al. 2007; Zhao et al. 2008; Fu et al. 2009). These QSARs hold when the mode of bioaccumulation is based on thermodynamic partitioning. When other processes are rate-limiting, such as with PFAS or surfactants, log K_{OW} based QSARs for the estimation of BCF fish data are not valid.

Further QSAR models are available, based on different descriptors and algorithms. The complexity of the required inputs and the function of the models can vary greatly, from simple hydrophobicity-driven approaches that consider biotransformation and diverse molecular parameters to mechanistically-based models. Each model has a specific applicability domain (AD), the compliance with which is crucial for prediction efficiency, as are other principles for predictions defined by the OECD (2023a) QSAR assessment framework (QAF). The efficiency of a prediction can only be evaluated for a specific substance or substance group and not generally for all QSAR models and all substances.

5.3.2 Estimation of BCF-QSAR values

Computational tools often include log K_{OW}-based methods as a worst-case approach, with some of them allowing for additional mitigating factors such as biotransformation, ionic interactions etc. Other methods use automated read-across based on different similarity measures. More recent methods rely on large databases for machine/deep learning using chemical structures and other features. Many computational tools are freely available, while numerous others are commercial or tailored in-house solutions, often based on confidential data. The intention of this part of the project was a case study to highlight potential issues that should be considered when estimated BCF-QSAR values are used, rather than to compare all existing models. Exemplarily, we used several freely available platforms providing predictive QSARs to estimate BCF for fish based on different methodological approaches.

5.3.2.1 QSAR tools for estimation of BCF fish data

The Estimations Programs Interface for Windows EPISuite (US EPA 2012) estimates the BCF of an organic compound using the compound's log K_{OW} based on Meylan et al. (1999) and has been expanded to include the estimation of the biotransformation rate (kM) in fish for estimation of the bioaccumulation factor (BAF) by the Arnot-Gobas method (Arnot and Gobas 2003). The model accounts for the mechanistic processes underlying bioconcentration and bioaccumulation, including chemical uptake from water across the gill surface (relevant to BCFs and BAFs) and from dietary sources (BAFs only), as well as chemical elimination through the gills, fecal egestion, growth dilution, and metabolic biotransformation (Arnot and Gobas, 2003). Additional processes considered in the calculations include bioavailability in the water column—since only the freely dissolved fraction can undergo bioconcentration—and absorption efficiencies at the gills and within the gastrointestinal tract. The model is not recommended for substances that ionise significantly, for pigments and dyes, or for perfluorinated compounds.

The software system ChemProp aims to predict compound properties from chemical structures (Schüürmann et al. 2007; UFZ 2021). Bioconcentration in fish can be calculated with single log K_{OW} relationships and BCF_{MAX} models, and complex log K_{OW} models such as the EUSES model. The

BCF models by Meylan et al. (1999) and by Arnot and Gobas (2003) are reimplementations of those available in EPISuite (US EPA 2012; Meylan et al. 1999; Arnot and Gobas 2003). The Read-Across BCF model is based on similarity of atom-centered fragments (ACF). The final result is a weighted average of different runs and includes a read-across code for coverage by the training set. AD compliance is checked (Kühne at al. 2009).

VEGAHUB is a platform of tools for the evaluation of chemical substances using in silico tools (IRCCS 2023). VEGA provides four BCF models: The CAESAR model is a hybrid model based on two separate computer-based models (Benfenati 2010). The BCF models by Meylan et al. (1999) and by Arnot and Gobas (2003) are re-implementations of those available in EPISuite (US EPA 2012; Meylan et al. 1999; Arnot and Gobas 2003). The k-nn-Read-Across model is a non-parametric model. This is the most different, compared to the previous models, because in the other models log K_{OW} has a main role (IRCCS 2023). VEGA provides an assessment of the reliability of the predictions and a visualization of the most similar compounds for read-across of the target substance. QSAR Model Reporting Format (QMRF) of VEGA models are available.

The Online Chemical Environment (OCHEM) is a web platform for data storage, model development and publishing of chemical information (Sushko et al. 2011). OCHEM includes a BCF QSAR model based on theoretical molecular descriptors (Gramatica and Papa 2005), and two models on the basis of the k-medoid algorithm using either DRAGON descriptors or EState indices and log K_{OW} (Brandmaier et al. 2012). The OCHEM output includes AD violation warnings.

The bioaccumulation assessment tool (BAT) is designed to collect, generate, evaluate and integrate various lines of evidence (LoE) (in silico, in vitro, in vivo) relevant for the B assessment into a consistent and transparent WoE that supports decision making (Armitage et al. 2021). It can also be used in data poor situations to predict generic lab fish and invertebrate BCF using chemical properties and estimates of biotransformation half-lives.

The BCFpro is a recently developed machine learning model trained on chemical properties and test conditions such as chemical concentration and exposure time on behalf of the German Environmental Agency (von der Ohe et al., personal communication). The (median of) BCF values for *Pimephales promelas*, *Onchorynhchus mykiss* and *Danio rerio* were only made available towards the end of the project, so these data were only included in some of the analyses.

The OECD QSAR Toolbox (OECD 2023b) is a computer software designed to make pragmatic qualitative and quantitative structure–activity relationship methods-based predictions of toxicity and other endpoints in a comprehensible and transparent manner (Schultz et al. 2018). The Toolbox provides information on chemicals in structure-searchable, standardised files to retrieve available experimental data and find data-rich analogues. Structural characteristics (alerts from profilers) and chemical properties can be used to group chemicals into a (toxicologically/mechanistically) meaningful category. In the *Data Gap Filling* module, the user can fill a data gap for their target substance using data from analogues with a trend analysis, read-across or existing QSAR models. In the *Report* section, users can generate a prediction report, a QMRF (QSAR model reporting format) and a justification document for the obtained prediction (*Read across justification document* – based on the RAAF principles (ECHA 2017c) and/or *QSAR justification document* – based on the QAF principles (OECD 2023a)).

Many more recent models of bioconcentration factors in fish have been described, but the algorithms are mostly proprietary and not publicly available. Examples are graph neural networks (Medina et al. 2021) and machine learning using topological and physicochemical properties (Kobayashi and Yoshida 2021), quantum-chemical descriptors (Fliszkiewicz and Sajdak 2023), in vitro/in vivo relationships (Xu et al. 2023), read-across (Pore et al. 2024) and molecular fingerprints (Dong et al. 2025). Explainable deep learning from chemical structures

(Zhao et al. 2022) and interpretable ensemble learning (Zhu et al. 2025) aim to address also mechanistic aspects.

According to their different approaches, the methods for estimation of BCF fish QSAR data have different strengths and weaknesses, which may lead to quite different quality of results for different classes of chemicals that may be within or outside the AD of the respective methods (Nendza et al. 2025).

5.3.2.2 Chemical datasets

BCF-QSAR values were estimated for two sets of case study chemicals. The "26 substances" dataset, with (almost) complete information on the bioaccumulation potential from alternative methods, is chemically diverse with molecular weights ranging from 166 to 500 g/mol, mean log K_{OW} from 2.12 to 7.68 and mean log K_{OA} from 4.13 to 15.3. The second dataset ("superhydrophobic") represent diverse chemical classes, such as POPs, PCB, PAH, flame retardants, pesticides, biocides, UV-filters, plasticisers, siloxanes and surfactants. The 33 case study chemicals all have consolidated (overall mean) log K_{OW} values greater than 6. The dataset "superhydrophobic" covers molecular weights from 227 to 971 g/mol, mean log K_{OW} from 5.97 to 10.7 and mean log K_{OA} from 5.08 to 17.9.

The two sets of case study chemicals as listed in the supplementary material [BCF QSAR; Table SI-1].

5.3.2.3 Predicted BCF-QSAR data

Estimated log BCF values for the two sets of case study chemicals were obtained using mostly freely available/public domain tools based on different conceptual methods: BCFBAF log BCF based on Meylan et al. (1999) and log BAF (upper trophic level) by the Arnot-Gobas method (Arnot and Gobas 2003) from EpiSuite (US EPA 2012), log BCF by tutorial MLRA, k-Medoid Dragon and k-Medoid PLS from OCHEM (Sushko et al. 2011), log BCF by k-nn Read-Across and the CAESAR model from VEGA (IRCCS 2023), log BCF by ACF Read-Across, the EUSES model and the log BCFmax (Nendza 1991) from ChemProp (UFZ 2021), log BCF generic lab fish from the BAT (Armitage et al. 2021), log BCF from xBCF (Zhao et al. 2022) and the median log BCF from BCFpro (von der Ohe et al., personal communication). This selection is exemplary and not exhaustive. Commercial software for BCF prediction was not used in this project.

In total, 296 and 319 individual log BCF values were acquired for the "26 substances" dataset and the "superhydrophobic" dataset, respectively. For most of the "26 substances", more than 10 estimates using different tools based on different conceptual methods were obtained. The only exception is GenX with only 8 calculated results. No QSAR predictions were recorded for substances that fall outside the AD of the respective models.

The predicted BCF data for the case study chemicals, their means across multiple methods (excluding the BCFmax worst-case model), their minimum and maximum values and their ranges, are provided in the supplementary material [BCF QSAR; Table SI-3].

Scatterplots of the estimates by the different methods are available in the supplementary material [BCF QSAR; Table SI-3] and illustrate the range and distribution of the predicted data.

5.3.3 Variability of BCF-QSAR values

Assessing the variability of BCF-QSAR values requires suitable reference values. The preferred benchmark for assessing the quality of BCF-QSAR estimates would be "gold standard" experimental data, for example from studies for regulatory use. As such data were not available within this project, data from the literature had to be used. Well known, a number of factors contribute to the experimental variability of measured BCF data, related to methodological,

biological and chemical factors (Weisbrod et al. 2007; Nendza et al. 2010; Wassenaar et al. 2020). Harmonised test protocols and detailed documentation can reduce some uncertainty, but the variability between data sources often exceeds 0.5 log units for multiple BCF values of the same substance. Notably, the BCF in fish is decisively influenced by the test species and two experimentally variable parameters: the applied concentration of the chemical in the external medium and the exposure time (von der Ohe et al., personal communication).

5.3.3.1 Experimental BCF fish data

For the "26 substances", literature was searched for experimental BCF values (see chapter 5.1). In addition, repositories of experimental BCF values are available in QSAR tools for the estimation of BCF fish data: EpiSuite (US EPA 2012) provides experimental data from its database. VEGA (IRCCS 2023) allows to extract the experimental data of the training sets of the implemented QSAR models (Meylan et al. 1999, Arnot and Gobas 2003, Benfenati 2010). Further experimental data were obtained from the database of the QSAR Toolbox (OECD 2023b). The collected data, their mean, median and maximum for the "26 substances" dataset are listed in Table 7. Details are provided in the supplementary material [BCF QSAR; Table SI-2]. No experimental BCF data were retrieved for the "superhydrophobic" substances.

The experimental BCF fish data set for the "26 substances" dataset includes 5 or more measured values for many of the substances. Their range is often within one log unit, thus mutually confirming each other. These data are therefore sufficiently plausible to be used as reference scale for QSAR estimates. Their individual data quality may not be good enough "for regulatory use", but their combined evidence as a benchmark for comparing predicted values by different QSAR models.

Table 7: Experimental log BCF fish data for the "26 substances" dataset collected from EpiSuite (US EPA 2012), VEGA (IRCCS 2023), the QSAR Toolbox (OECD 2023b) and a literature search (chapter 5.1); for more details see the supplementary material [BCF QSAR; Table SI-2].

Substance name	CAS	Consolid. log K_{OW}	EpiSuite	VEGA k-nn	VEGA Caesar	VEGA Arnot Gobas	VEGA Meylan	OECD QSAR Toolbox	Literature screening	Exp. MEAN log BCF	Exp. Median log BCF	Exp. Max (worst case) log BCF
Simazine	122-34-9	2.12	0.56	1.11		1.11	0.56		0-2.33	1.07	1.11	2.33
Azoxystrobin	131860-33-8	3.00										
GenX	62037-80-3	3.26							0-0.9	0.65	0.65	0.90
Terbutryn	886-50-0	3.41							0.95-1.64	1.30	1.30	1.64
17a-Ethinylestradiol	57-63-6	3.84				2.92		2.43 - 2.52		2.70	2.70	2.92
β-Hexachloro-cyclohexane	319-85-7	3.87	3.25	2.56	2.77	2.99	3.06	2.47	3.02-3.16	2.88	2.99	3.25
Prochloraz	67747-09-5	3.91		2.59					2.29-2.59	2.53	2.53	2.59
Fluorene	86-73-7	4.06	2.72	2.81	2.78	2.88	2.72	1.76 - 2.86	3.11	2.76	2.78	3.11
Fluoxetine	54910-89-3	4.33							1.11-2.52	1.81	1.81	2.52.
Diclofenac	15307-86-5	4.35							0.7-0.95	0.83	0.83	0.95
Anthracene	120-12-7	4.41	3.26	3.25	2.99	3.14	3.26		2.28-3.78	3.23	3.25	3.78

Substance name	CAS	Consolid. log K_{OW}	EpiSuite	VEGA k-nn	VEGA Caesar	VEGA Arnot Gobas	VEGA Meylan	OECD QSAR Toolbox	Literature screening	Exp. MEAN log BCF	Exp. Median log BCF	Exp. Max (worst case) log BCF
PFOS	1763-23-1	4.42		3.73	3.73	3.47			2.41-3.73	3.59	3.60	3.73
Phenanthrene	85-01-8	4.43	3.40	3.31	3.01	3.28	3.40		2.85 - 3.79	3.32	3.36	3.79
Pentachlorophenol	87-86-5	4.81	2.66	2.67	2.50	2.24	2.66	2.29 - 2.68		2.54	2.58	2.68
Lauric Acid	143-07-7	4.84	2.40			2.41	2.40		2.41	2.40	2.40	2.41
Chlorpyrifos	2921-88-2	4.89	3.18	3.01		2.74	3.18	3.14	1.6-3.5	2.97	3.07	3.50
Triclosan	3380-34-5	4.91	1.72	1.67	1.72	1.54	1.72	0 - 3.3		1.67	1.69	3.30
Pyrene	129-00-0	4.91	3.18	2.75	2.56	2.85	3.18	1.7 - 2.99	1.36-3.13	2.73	2.75	3.18
Octamethylcyclo-tetrasiloxane (D4)	556-67-2	4.98		4.09				3.70	4.09-4.17	3.97	4.09	4.17
Trifloxystrobin	141517-21-7	5.08		2.63					2.57-2.73	2.64	2.64	2.73
Pentachlorobenzene	608-93-5	5.11		3.93	3.49	3.82	3.75	3.6 - 4.36	3.88-3.93	3.81	3.86	4.36
Methoxychlor	72-43-5	5.25					2.50	2.82 - 3.92	2.79-3.33	2.98	3.06	3.92
o-Terphenyl	84-15-1	5.58		3.11		3.11	3.48		2.7-4.11	3.28	3.26	4.11
Benzo[a]pyrene	50-32-8	6.07		2.73	2.69	3.10		0.48 - 2.56	2.58-2.8	2.55	2.69	3.10

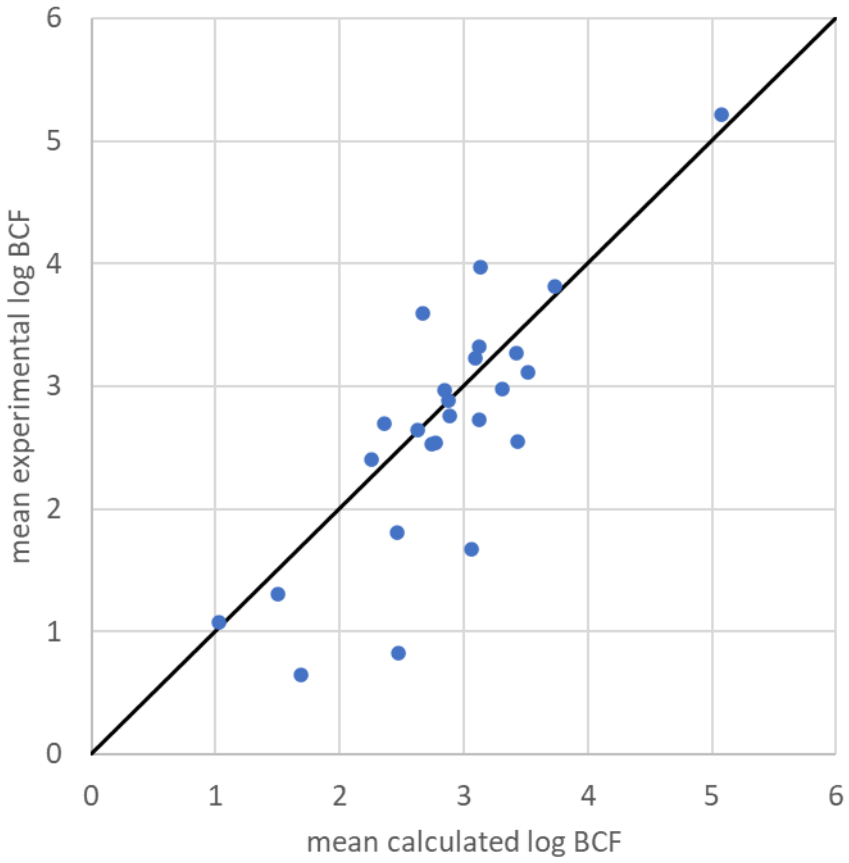
Substance name	CAS	Consolid. log K_{OW}	EpiSuite	VEGA k-nn	VEGA Caesar	VEGA Arnot Gobas	VEGA Meylan	OECD QSAR Toolbox	Literature screening	Exp. MEAN log BCF	Exp. Median log BCF	Exp. Max (worst case) log BCF
PCB 153	35065-27-1	7.08	5.43	5.69		4.59	5.43		3.95 - 5.87	5.21	5.43	5.87
UV-234	70321-86-7	7.68							3.11	3.11	3.11	3.11

Colour Coding: red: B (BCF >2000), dark red: vB (BCF >5000)

5.3.3.2 Comparative analyses of BCF-QSAR estimates for the 26 substances

The predicted BCF fish QSAR data of the "26 substances" dataset [BCF QSAR; Table SI-3] are variable with ranges exceeding one log unit and more across methods. The data distribution does not reveal any systematic patterns, with different outliers to different extends. Comparing the mean of the predicted BCF-QSAR data per substances and the mean of their experimental data reveals agreement mostly within half a log unit with r^2 of 0.67 (Figure 4). The largest negative deviations (estimate much lower than measured value "false negative" (data points above the 1:1 line)) relate to PFOS and octamethylcyclotetrasiloxane (D4), while the largest positive deviations (estimate much higher than measured value "false positive" (data points below the 1:1 line)) concern ionisable substances such as diclofenac, fluoxetine, triclosan and GenX. If these six substances are excluded from the correlation, r^2 increases to 0.89.

Figure 4: Comparison of the mean of the BCF-QSAR estimates for the "26 substances" with the mean of their experimental BCF fish data. The solid line is the 1:1 line.



Source: own illustration, AL-Luhnstedt

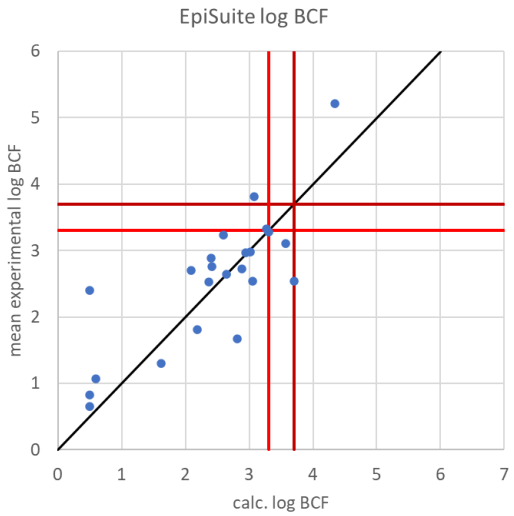
A closer examination of the BCF-QSAR estimates according to different methods requires consideration of the appropriate reference scale. Using the mean of the experimental BCF data is the simplest option. From a statistical point of view, the median would be preferable, as the data are not normally distributed. However, the distinction is negligible here, as the mean and median differ only marginally in this dataset (Table 7). Using the highest available experimental BCF data as a "worst-case" scenario may by appealing from an environmental protection perspective but should be rejected because the data search may not have been exhaustive enough to identify the maximum values for all substances, i.e. this reference scale is likely inconsistent.

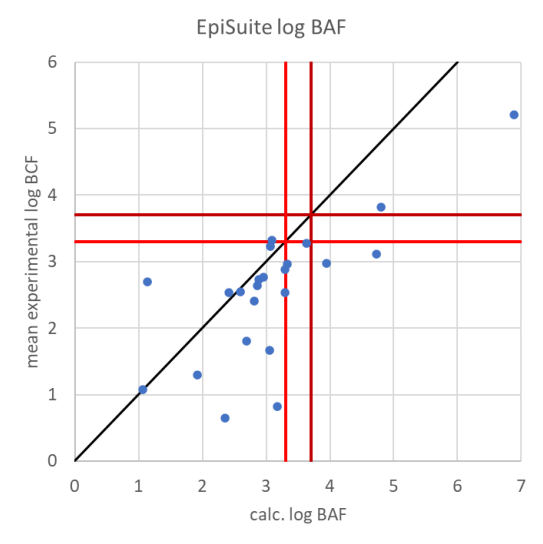
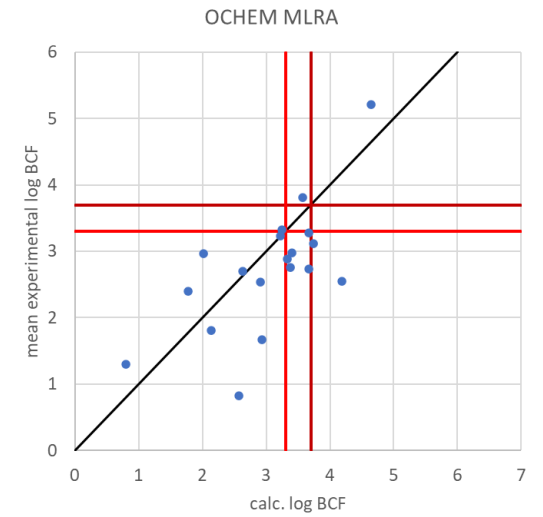
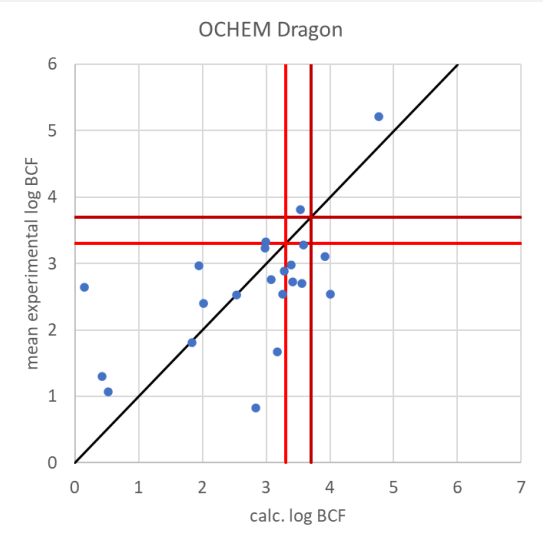
Table 8 presents the scatterplots of the BCF-QSAR estimates by 13 different methods and their mean relative to the mean of the experimental BCF fish data. Please note, PFOS and D4 are excluded from the correlations presented in Table 8. For a comprehensive assessment of the analysed data, the thresholds for B (BCF > 2000, light red) and vB (BCF > 5000, dark red) are shown. For orientation: The data points in the lower left and in the upper right quadrant are correctly classified as nonB or B/vB, respectively. Those in the lower right quadrant are overprotective "false positive" estimates. "False negative" predictions in the upper left quadrant, which are incorrectly classified as nonB instead of B/vB are problematic. Encouragingly, we observe only one substance with a clear "false negative" prediction, namely pentachlorobenzene from EpiSuite log BCF.

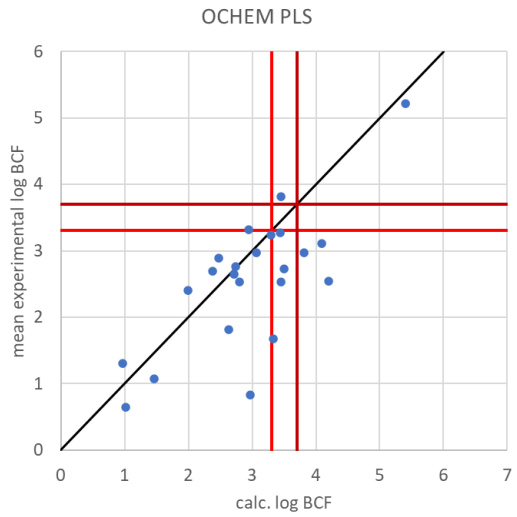
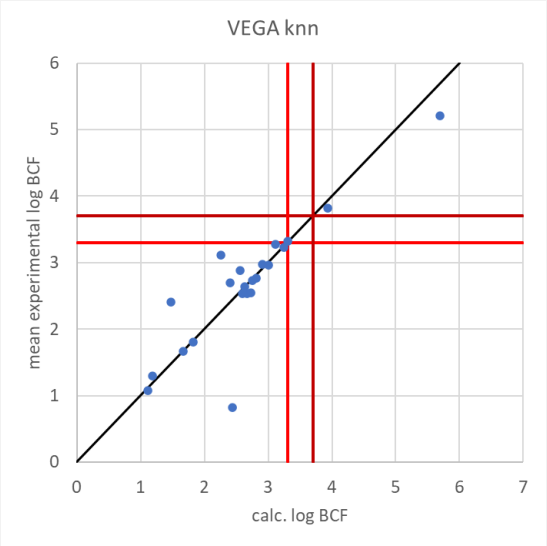
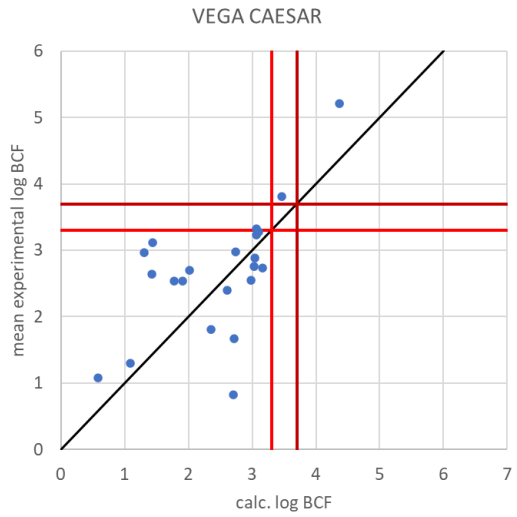
The performance of various methods for calculating log BCF fish QSAR data of the "26 substances" dataset is rather disappointing according to their correlation coefficients. Ten of the 14 models have r^2 less than 0.6. Perhaps surprisingly, the log K_{OW}-based EpiSuite log BCF according to Meylan et al. (1999), is among the top four models with r^2 of 0.64, although its performance is hampered by the above-mentioned "false negative" prediction. Among the top three models are a read-across model (VEGA k-nn (IRCCS 2023) with r^2 of 0.77) and a deep learning model (xBCF (Zhao et al. 2022) with r^2 of 0.82). A consensus model, i.e. the mean value of all log BCF fish QSAR data predicted in this study (except BCF max), also performs remarkably well with r^2 of 0.74.

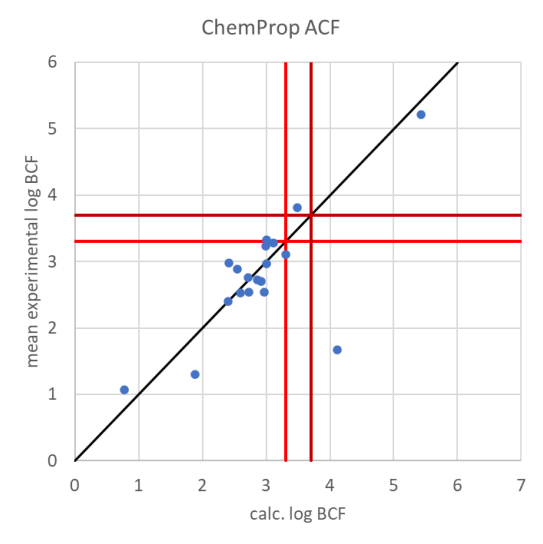
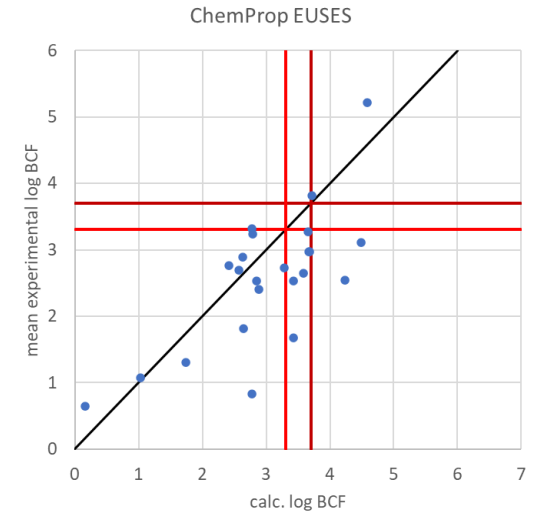
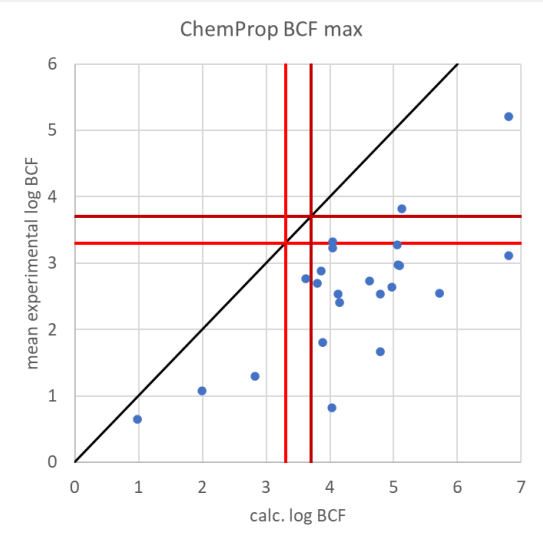
The corresponding plots and statistics for the median and the maximum (worst-case) of the available experimental BCF fish data are included in the supplementary material [BCF QSAR; Table SI-2]. Compared to the scatterplots in Table 8, shifting the scale of the y-axes to higher values while keeping the data distribution on the x-axis unchanged, leads to increased rates of "false negative" predictions.

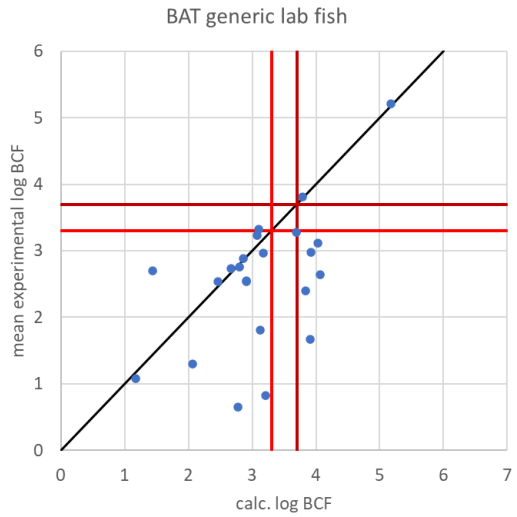
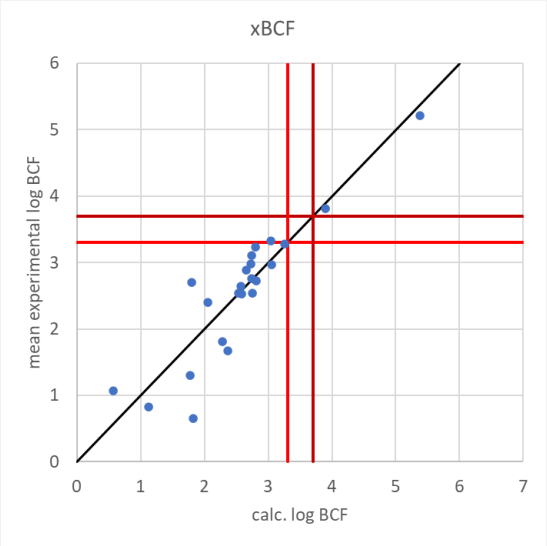
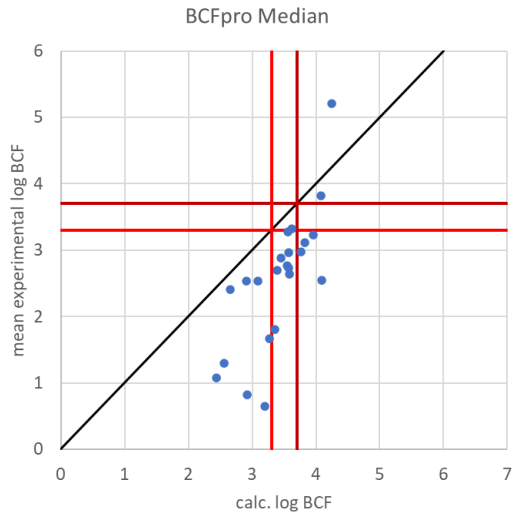
Table 8: Performance of different methods for calculating log BCF fish QSAR data of the "26 substances" dataset (excluding PFOS and D4). The solid black line is the 1:1 line. The light red lines represent the threshold for B (BCF > 2000) and the dark red lines represent the threshold for vB (BCF > 5000).

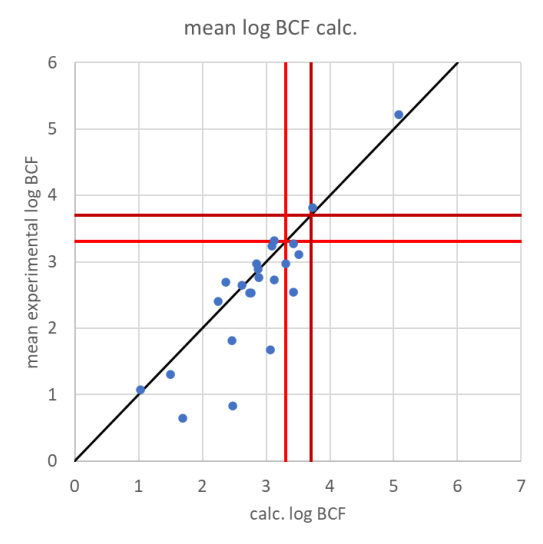
Tool	Method	r^2	
EPISuite (US EPA 2012)	log BCF based on Meylan et al. (1999)	0.64	

Tool	Method	r^2	
	log BAF (upper trophic level) by the Arnot-Gobas method (Arnot and Gobas 2003)	0.54	
OCHEM (Sushko et al. 2011)	log BCF by tutorial MLRA	0.46	
	log BCF by k-Medoid Dragon	0.37	

Tool	Method	r^2	
	log BCF by k-Medoid PLS	0.56	
VEGA (IRCCS 2023)	log BCF by k-nn Read-Across	0.77	
	log BCF by the CAESAR model	0.37	

Tool	Method	r^2	
ChemProp (UFZ 2021)	log BCF by ACF Read-Across	0.57	
	log BCF by the EUSES model	0.52	
	log BCFmax based on Nendza (1991)	0.54	

Tool	Method	r^2	
BAT (Armitage et al. 2021)	log BCF Generic lab fish	0.31	
xBCF (Zhao et al. 2022)	Deep Learning log BCF	0.82	
BCFpro (von der Ohe et al., personal communication)	Machine Learning log BCF	0.57	

Tool	Method	r^2	
This study	Mean of predicted log BCF fish QSAR data (except BCF max)	0.74	

Source: own illustration, AL-Luhnstedt

5.3.3.3 Comparative analyses of BCF-QSAR estimates for "superhydrophobic" substances

The comparative analyses of BCF-QSAR estimates for "superhydrophobic" substances are limited compared to the "26 substances" as only predictions from EpiSuite (US EPA 2012), OCHEM (Sushko et al. 2011), VEGA (IRCCS 2023) and ChemProp (UFZ 2021) were available. Furthermore, no experimental BCF data were obtained as reference values.

The predicted BCF fish QSAR data for the "superhydrophobic" substances [BCF QSAR; Table SI-3] are variable with ranges exceeding one log unit and more across methods. Their mean predicted BCF value is above 5000 for seven substances, indicating vB properties, namely for Mirex, DDT, 4,4'-DDD, 4,4'-DDE, PCB 77, PCB 142 and PCB 153. Another six substances have mean predicted BCF between 2000 and 5000, namely benzo[a]pyrene, hexabromobenzene, 2,3,4,5-tetrabromo-6-chlorotoluene, difenacoum, UV-234 and 2,4,6-tri-tert-butylphenol. Among the substances with mean predicted BCF less than 2000 are, for example, surfactants such as hexadecylamine, N-methylhexadecylamine, N,N-dimethyltridecylamine, N,N-dimethyltetradecylamine, N,N-dimethylhexadecylamine and C18 benzalkonium cation.

The BCF-QSAR estimates for two substances in the "superhydrophobic" dataset show particularly large ranges of more than 5 log units [BCF QSAR; Table SI-3]: Predictions of more than 5000 (vB) are observed for dodecamethylpentasiloxane (L5) and dodecamethylcyclohexasiloxane (D6) with BCFBAF log BAF (upper trophic level) by the Arnot-Gobas method (Arnot and Gobas 2003) from EpiSuite (US EPA 2012), log BCF by k-nn Read-Across from VEGA (IRCCS 2023), log BCF by the EUSES model and the log BCFmax (Nendza 1991) from ChemProp (UFZ 2021). Much lower predictions of less than 2000 (nonB) result from BCFBAF log BCF based on Meylan et al. (1999) from EpiSuite (US EPA 2012), log BCF by tutorial MLRA, k-Medoid Dragon and k-Medoid PLS from OCHEM (Sushko et al. 2011), log BCF by the CAESAR model from VEGA (IRCCS 2023) and log BCF by ACF Read-Across from ChemProp (UFZ 2021). The observed discrepancies mean that many methods are not suitable for these siloxanes. They are likely not covered by the QSARs' ADs, even though no AD violation warnings were issued.

5.3.3.4 Comparative analyses of BCF-QSAR methods

Several evaluations can be used to illustrate the variability of calculated log BCF values obtained by different methods. The variability of the individual methods can be displayed relative to the consolidated log K_{OW} (see plots below the data table in [BCF QSAR; Table SI-3]. The scatter plots (Table 8) reveal the performance of different methods for calculating log BCF fish QSAR data in comparison to the mean experimental BCFs. A correlation matrix (Table 9) can demonstrate the similarities or dissimilarities between the results from different methods. Statistics such as root mean square error (RMSE) and maximal deviations in relation to, for example, the overall mean log BCF value can address the consistency of BCF-QSAR predictions (Table 10 and Table 11).

The comparative analyses of BCF-QSARs may first discriminate the underlying methodologies: Three principally different methods are available for estimating BCF values using the computational tools applied in this study: The most widely used are log K_{OW}-based methods, which are considered a worst-case approach, although some of them take additional mitigating factors such as biotransformation, ionic interactions, etc. into account. Alternatively, other methods use automated read-across based on different similarity measures to extrapolate the BCF from one or several source substances to the target chemical. More recent methods rely on large databases for machine/deep learning using various encodings of chemical structures and other features.

The correlation matrix (Table 9) reveals that the estimates by the different tools are only partially correlated, in most cases the regression coefficient r is less than 0.8. Higher levels of agreement are seen between the log K_{OW}-based methods such as BCFBAF log BCF based on Meylan et al. (1999), the EUSES model and the log BCFmax (Nendza 1991) from ChemProp (UFZ 2021). Methodological overlap may be indicated by high correlation between log BAF (upper trophic level) by the Arnot-Gobas method (Arnot and Gobas 2003) and the log BCF generic lab fish from the BAT (Armitage et al. 2021). Also, the tools for automated read-across, i.e. the log BCF by k-nn read-across from VEGA (IRCCS 2023) and the log BCF by ACF read-across from ChemProp (UFZ 2021), are well correlated. High correlation coefficients are seen as well among the OCHEM results (Gramatica and Papa 2005, Brandmaier et al. 2012).

As was to be expected, the regression coefficients of most methods with the mean calculated BCF data are between 0.8 and 0.9. As already seen in the scatter plots (Table 8), the highest correlations with the mean experimental BCFs are confirmed for EpiSuite log BCF according to Meylan et al. (1999), the VEGA k-nn read-across model (IRCCS 2023), the deep learning model xBCF (Zhao et al. 2022) and the consensus model, i.e. the mean value of all log BCF fish QSAR data predicted in this study (except BCF max).

Table 10 and Table 11 summarise the performance of the different methods in terms of availability (number (n) of estimates obtained), accuracy (root mean square error (RMSE) relative to the overall mean calculated log BCF value and variability represented by the scatter, i.e. largest deviations in either direction (Δ MIN, Δ MAX) for the “26 substances” and “superhydrophobic” datasets. With RMSEs mostly below 1, the estimates for the “26 substances” are more consistent than the predictions for the “superhydrophobic” chemicals, with more RMSEs above 1.

Overall, the comparative analyses of BCF-QSAR methods revealed that no systematic pattern is evident. No method is consistently superior (or inferior) for the case study chemicals.

Table 9: Correlation matrix (regression coefficients r) of the estimates by different methods for calculating log BCF fish QSAR data of the "26 substances" dataset (excluding PFOS and D4), the mean calculated and experimental log BCF values, and the consolidated log K_{OW}.

	Epi-Suite log BCF	Epi-Suite log BAF	OCHEM MLRA	OCHEM Dragon	OCHEM PLS	VEGA k-nn	VEGA CAESAR	ChemProp ACF	Chem-Prop EUSES	Chem-Prop BCF max	BAT gen lab fish	xBCF	BCFpro	Mean log BCF calc.	Mean log BCF exp.	log K_{OW}
Epi-Suite log BCF	1															
Epi-Suite log BAF	0.63	1														
OCHEM MLRA	0.75	0.61	1													
OCHEM Dragon	0.65	0.61	**0.95**	1												
OCHEM PLS	**0.82**	0.79	**0.87**	0.79	1											
VEGA knn	0.69	0.77	0.72	0.62	0.76	1										
VEGA CAESAR	0.48	0.64	0.71	0.76	0.66	0.68	1									
Chem-Prop ACF	0.74	0.77	0.58	0.75	0.77	0.75	0.67	1								
Chem-Prop EUSES	**0.81**	0.68	0.71	0.66	**0.88**	0.62	0.48	0.75	1							
Chem-Prop	**0.81**	0.73	0.69	0.66	**0.89**	0.62	0.45	0.77	**0.99**	1						

	Epi-Suite log BCF	Epi-Suite log BAF	OCHEM MLRA	OCHEM Dragon	OCHEM PLS	VEGA k-nn	VEGA CAESAR	ChemProp ACF	Chem-Prop EUSES	Chem-Prop BCF max	BAT gen lab fish	xBCF	BCFpro	Mean log BCF calc.	Mean log BCF exp.	log K_{ow}
BCF max																
BAT gen lab fish	0.51	**0.87**	0.47	0.42	0.68	0.56	0.58	0.76	0.68	0.70	1					
xBCF	**0.82**	**0.85**	0.67	0.62	0.76	**0.88**	0.69	**0.85**	0.69	0.71	0.72	1				
BCFpro	**0.80**	0.65	**0.82**	0.66	0.76	0.78	0.63	0.67	0.66	0.66	0.56	0.79	1			
Mean log BCF calc.	**0.85**	**0.88**	**0.88**	**0.82**	**0.93**	**0.84**	0.79	**0.87**	**0.84**	**0.85**	0.77	**0.91**	**0.85**	1		
Mean log BCF exp.	0.79	0.75	0.68	0.66	0.75	0.79	0.67	0.75	0.71	0.72	0.60	**0.88**	0.75	**0.87**	1	
log Kow	0.75	0.78	0.64	0.65	**0.83**	0.54	0.47	0.72	**0.89**	**0.92**	0.79	0.71	0.69	**0.84**	0.69	1

Table 10: Comparison of the performance of different methods for calculating log BCF fish QSAR data of the "26 substances" dataset (availability: number of data points (n) obtained, accuracy: root mean square error (RMSE), variability: largest negative deviation (Δ MIN) underestimating and largest positive deviation (Δ MAX) overestimating log BCF in relation to the overall mean log BCF value).

Tool	Method	n	RMSE	Δ MIN	Δ MAX
EPISuite (US EPA 2012)	log BCF based on Meylan et al. (1999)	26	0.84	-2.31	0.88
	log BAF (upper trophic level) by the Arnot-Gobas method (Arnot and Gobas 2003)	26	0.86	-1.26	3.02
OCHEM (Sushko et al. 2011)	log BCF by tutorial MLRA	19	0.43	-0.95	0.62
	log BCF by k-Medoid Dragon	24	0.81	-2.61	1.16
	log BCF by k-Medoid PLS	26	0.69	-2.58	0.64
VEGA (IRCCS 2023)	log BCF by k-nn Read-Across	25	0.62	-1.52	0.92
	log BCF by the CAESAR model	25	0.99	-2.76	0.23
ChemProp (UFZ 2021)	log BCF by ACF read-across	21	0.51	-1.01	0.93
	log BCF by the EUSES model	26	0.47	-1.21	0.82
	log BCFmax based on Nendza (1991)	26	1.56	-0.39	3.04
BAT (Armitage et al. 2021)	log BCF generic lab fish	26	0.63	-0.96	1.46
xBCF (Zhao et al. 2022)	deep learning log BCF	26	0.57	-1.45	1.00

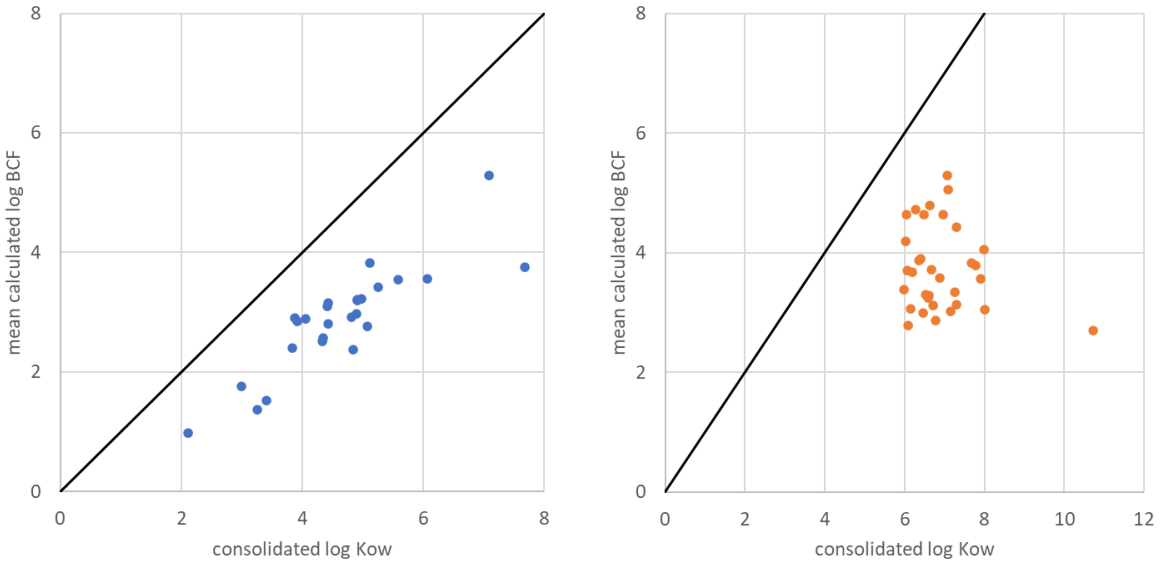
Table 11. Comparison of the performance of different methods for calculating log BCF fish QSAR data of the "superhydrophobic" dataset (availability: number of data points (n) obtained, accuracy: root mean square error (RMSE), variability: largest negative deviation (Δ MIN) underestimating and largest positive deviation (Δ MAX) overestimating log BCF in relation to the overall mean log BCF value).

Tool	Method	n	RMSE	Δ MIN	Δ MAX
EPISuite (US EPA 2012)	log BCF based on Meylan et al. (1999)	33	0.65	-1.83	0.92
	log BAF (upper trophic level) by the Arnot-Gobas method (Arnot and Gobas 2003)	33	1.41	-1.64	3.03
OCHEM (Sushko et al. 2011)	log BCF by tutorial MLRA	32	1.04	-2.98	2.46
	log BCF by k-Medoid Dragon	32	0.71	-1.50	2.73
	log BCF by k-Medoid PLS	33	1.02	-2.39	3.57
VEGA (IRCCS 2023)	log BCF by k-nn Read-Across	32	1.12	-2.08	1.22
	log BCF by the CAESAR model	33	1.62	-2.90	0.03
ChemProp (UFZ 2021)	log BCF by ACF read-across	25	0.77	-1.87	0.27
	log BCF by the EUSES model	33	1.09	-3.47	1.79
	log BCFmax based on Nendza (1991)	33	2.56	1.18	3.80

5.3.3.5 Comparative analyses of BCF-QSAR estimates relative to log K_{OW}

Underlying trends are revealed by comparing the mean of the calculated log BCF fish QSAR data and consolidated log K_{OW} for the "26 substances" dataset and the "superhydrophobic" dataset in Figure 5. Increasing log BCF with increasing log K_{OW} between 1 and 6 is clearly evident for the "26 substances" dataset, while the postulated hydrophobicity cut-off above log K_{OW} 6 is not confirmed for the "superhydrophobic" dataset. Rather, a levelling of log BCF can be observed, which scatters between 3 and 5. The apparent loss in linear relationships for superhydrophobic compounds has been attributed – in part – to experimental artefacts (Jonker and van der Haijden 2007a, b). Indeed, many earlier BCF values that were "too low" relative to the log K_{OW} of the test substances have been revised upwards in more recent tests. Relevant factors are bioavailability (e.g. freely dissolved exposure concentrations, sorption, limits of water solubility, effects of solvents or other solubilisers) and kinetics (e.g. not reaching steady-state, reduced diffusion in water phases, effects of size and lifestage of test organisms).

Figure 5: Relationship between the mean of the calculated log BCF fish QSAR data and consolidated log K_{OW} for the "26 substances" dataset (left) and the "superhydrophobic" dataset (right).



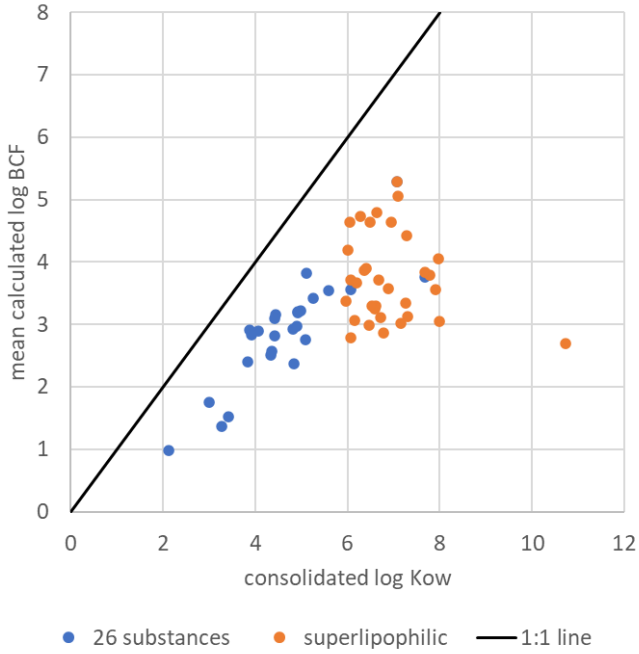
Source: own illustration, AL-Luhnstedt

Theoretical considerations substantiate curvilinear BCF/hydrophobicity relationships, though for chemicals with much higher log K_{OW} ~10, due to aqueous phase diffusion control of water to lipid transfer, differences in phase (solvent) properties of natural lipids and octanol, influence of steric conformations and differences in thermodynamic properties of partitioning, e.g. enthalpy changes (Gobas et al. 2006; Müller and Nendza 2007; Jonker and van der Haijden 2007b).

Large superhydrophobic compounds also permeate through membranes, but at much slower rates, due to their low solubility in water layers. Hence, uptake phases in experimental tests would have to be extended in time. Reduced uptake of large molecules is not substantiated. Data on the oral absorption of antibiotics indicate that even large substances with a high molecular weight (MW > 1000 g/mol) can pass through membranes and be absorbed into organisms (Gobas et al. 2006; Seydel and Wiese 2002). The slow rate of elimination gives high log K_{OW} chemicals their inherent bioaccumulative potential (Arnot and Gobas 2003).

Combining the two sets of case study chemicals (Figure 6) provides a more comprehensive overview of the relationship between overall mean calculated log BCF fish QSAR data and the consolidated log K_{OW} for the "26 substances" (blue) and the "superhydrophobic" (orange) dataset. For data and details see the supplementary material [BCF QSAR; Table SI-3].

Figure 6: Relationship between the mean of the calculated log BCF fish QSAR data and consolidated log K_{OW} for the "26 substances" (blue) and the "superhydrophobic" (orange)" dataset.



Source: own illustration, AL-Luhnstedt

5.3.4 Consolidation of BCF-QSAR values

Based on the case study with the "26 substances" and the "superhydrophobic" dataset, it was demonstrated that there are several reasons and issues that can lead to variability and uncertainty in BCF-QSAR predictions. These have varying relevance and consequences depending on the chemical substance concerned, the extent of available data on related endpoints, and the regulatory context.

Data quality requirements depend on the situation to be considered: When a precise BCF estimate is needed, e.g. close to a regulatory threshold, high accuracy and low uncertainty are required. If other available information suggests that a BCF far from regulatory thresholds is to be expected, a reduced approach that allows for inaccuracies may be sufficient.

In both situations, the ideal solution would be a high quality QSAR, such as a model based on rate-limiting processes, e.g. lipid partitioning and metabolism for a series of congeners, or a read-across model with sufficiently close analogues in its database. Then, the target compound is covered by the model's AD resulting in high prediction confidence.

Under less perfect circumstances, consensus modelling with multiple independent QSARs may be an option to reduce uncertainties to the level needed to come to a robust conclusion about the bioaccumulation potential of the target chemical. Notably, independent QSARs are models based on different conceptual methodologies, e.g. one log K_{OW}-based model (with or without mitigating factors), one model using automated read-across, one mechanistically-based (ADME) model, one model based on deep learning. How many of these, or even more, QSARs may be needed for consolidation of BCF-QSAR values depends on the level of acceptable uncertainty, as outlined above.

Using several models based on the same conceptual methodology offers little additional value of information and may rather disguise some statistical variation. This makes them insufficient for

consensus modelling, i.e. a WoE or an averaging approach, to reduce uncertainties and consolidate variable, possibly conflicting, results. Combining multiple estimates from several independent methods, instead, takes advantage of their strength while compensating outliers (Lombardo et al. 2014; Nendza et al. 2013). As confirmed by the present data analyses, the overall mean of the calculated log BCF values is well correlated with the available experimental log BCF data for the "26 substances" dataset.

5.3.5 Conclusions

A case study was carried out as part of this study having three main objectives: (1) to estimate multiple BCF fish QSAR data by different methods for diverse chemicals, (2) to analyse the variability of BCF fish QSAR estimates regarding the underlying methods, and (3) to recommend approaches to obtain reliable and robust BCF fish QSAR estimates for the hazard and risk assessment of chemicals.

The systematic data analyses of this study indicate that no QSAR methodology is per se superior for predicting QSAR-BCF estimates. Among the top four models in this study are the log K_{OW}-based EpiSuite according to Meylan et al. (1999), the VEGA k-nn read-across model (IRCCS 2023), the deep learning model xBCF (Zhao et al. 2022) and a consensus model, i.e. the mean value of all log BCF fish QSAR data predicted in this study.

The reliability of individual QSAR predictions should be evaluated in the context of additional information, considering, for example, read-across from similar substances, evidence derived from log K_{OW} and biohalf-life values and in vitro assays.

If two pieces of information are inconsistent, for example if a predicted and an experimental BCF value are disagreeing, analyse both, the reliability of the QSAR prediction and the experimental study for possible shortcomings.

Consensus modelling for consolidation of BCF values combines estimates from several independent methods, reduces uncertainties and takes advantage of their strength while compensating outliers (Lombardo et al. 2014; Nendza et al. 2013). As confirmed by the present data analyses, the overall mean of the calculated log BCF values is well correlated with the available experimental log BCF data for the "26 substances" dataset.

5.4 Physicochemical parameters for bioaccumulation assessment: log K_{OW}

Hydrophobicity is a key parameter for assessing potential environmental impacts of chemicals. This property determines the preference of a substance for either aqueous or non-aqueous phases and, thus, the rate and direction of transport of chemicals in the environment. Partitioning and hydrophobic bonding to target sites drives the (eco)toxicity of contaminants and relates to their accumulation potential in soil, sediment and biota. The unitless equilibrium partition coefficient K_{OW} between the two immiscible phases 1-octanol and water (Equation 1) is used to measure hydrophobicity (Nendza et al. 2025):

$$K_{OW} = \frac{\text{equilibrium concentration of chemical in 1-octanol (mol/L)}}{\text{equilibrium concentration of chemical in water (mol/L)}} \qquad \text{Eq. 1}$$

5.4.1 Estimation of log K_{OW} values

Methods to determine the log K_{OW} in regulatory acceptable ways are outlined, for example, in ECHA Guidance Documents (ECHA 2017b; ECHA 2023a). The default experimental method is the shake-flask method (OECD 1995), which works well for organic substances with intermediate hydrophobicity and substantial water solubility (log K_{OW} range -2 to 4). The generator column method (EPA OPPTS 830.7560 (US EPA 1996)) is suitable for more hydrophobic chemicals (log

K_{OW} range 1 to 6). The slow-stirring method (OECD 2022a) was developed for highly hydrophobic substances (log K_{OW} > 4.5 up to 8.2). Experimental determination of log K_{OW} by equilibrium partitioning usually provides reliable results with a repeatability of ± 0.3 log units, though higher variability exceeding 1 log unit has been reported (Nendza et al. 2010).

Chromatographic methods, such as OECD TG 117 (OECD 2022b), differ by using a dynamic process, which may be more similar to the distributions in environmental systems. Applicable for non-ionogenic substances with log K_{OW} in the range of 0 to 6, the accepted variation is ± 0.5 log units. Due to the limitations of the OECD TG 117 HPLC derived log K_{OW} data, the ECHA Guidance (ECHA 2023a) calls to support these with QSAR generated estimates, especially if the log K_{OW} is in the range of one log unit below or above the screening value of log K_{OW} = 4.5 (Nendza et al. 2025).

The universal relevance of partitioning in the assessment of chemicals led to early attempts to estimate it directly from the chemical structures (Hansch & Fujita 1964; Rekker 1977; Hansch & Leo 1979). Starting with the assumption of an additive, constitutive nature of partition coefficients, Rekker (1977) and Hansch and Leo (1979) published the first group contribution methods in which the log K_{OW} of a compound is the sum of the hydrophobicity contributions of its constituents. The structures are virtually decomposed into substructural fragments (f_i) with individual contributions (a_i) to the hydrophobicity of the entire molecules. Additionally, correction terms (b_i) account for interactions between different fragments in the molecules (F_i) (Nendza et al. 2025):

$$\log K_{OW} = \Sigma\, a_i\, f_i + \Sigma\, b_i\, F_i \qquad \text{Eq. 2}$$

Many log K_{OW} estimation tools have since been derived based on group contributions, which cover most organic compounds, e.g. (US EPA 2012; Data Warrior 2013; UFZ 2021; US EPA 2022; ACD/Labs 2023; ARC 2023; Chemaxon 2023; IRCCS 2023; Molinspiration 2023; PerkinElmer Informatics 2023). The methods differ in the underlying datasets and in the definition of the particular fragments and factors (Nendza et al. 2025).

Alternative techniques for log K_{OW} estimation include linear solvation energy relationships (LSER), e.g. (Kamlet et al. 1988; Endo & Goss 2014; Ulrich et al. 2017), automated read-across (k-nn), e.g. (Mansouri et al. 2018; UFZ 2021; IRCCS 2023), deep learning neural networks, e.g. (Sushko et al. 2011; Ulrich et al. 2021; Zhao et al. 2022), quantum chemistry-based methods, e.g. (Klamt & Schüürmann 1993; Glüge et al. 2022; BIOVIA 2023), and consensus modelling, e.g. (UFZ 2021; ACD/Labs 2023; ARC 2023).

5.4.2 Variability of log K_{OW} values

According to their different approaches, the methods for log K_{OW} estimation have different strengths and weaknesses, which may lead to quite different quality of results for different classes of chemicals that may be within or outside the AD of the respective methods. Knowing the accuracy, reliability and variability of log K_{OA} estimates is especially important when substances are assessed against (regulatory) thresholds based on log K_{OW} (Nendza et al. 2025).

The systematic comparison of log K_{OW} values, measured and calculated, has revealed that deviations of more than one log unit are the rule rather than the exception (Figure 7). Outliers occur with all methods and tools, regardless of the substance class. No method is throughout superior (or inferior) for estimating the log K_{OW} of the diverse case study chemicals (Table 12). To decide whether a log K_{OW} value is correct or an outlier, consolidation by further independent measurements and/or estimates is needed (Nendza et al. 2025).

Figure 7: **Variability of log K_{OW} estimates for 173 case study chemicals obtained by different experimental and computational methods. The x-axis represents the 173 case study chemicals sorted by increasing mean log K_{OW}, the y-axis reveals the range of the log K_{OW} data for each chemical. See Table 12 for details on the performance of the tools and references. Reproduced from Nendza et al. 2025**

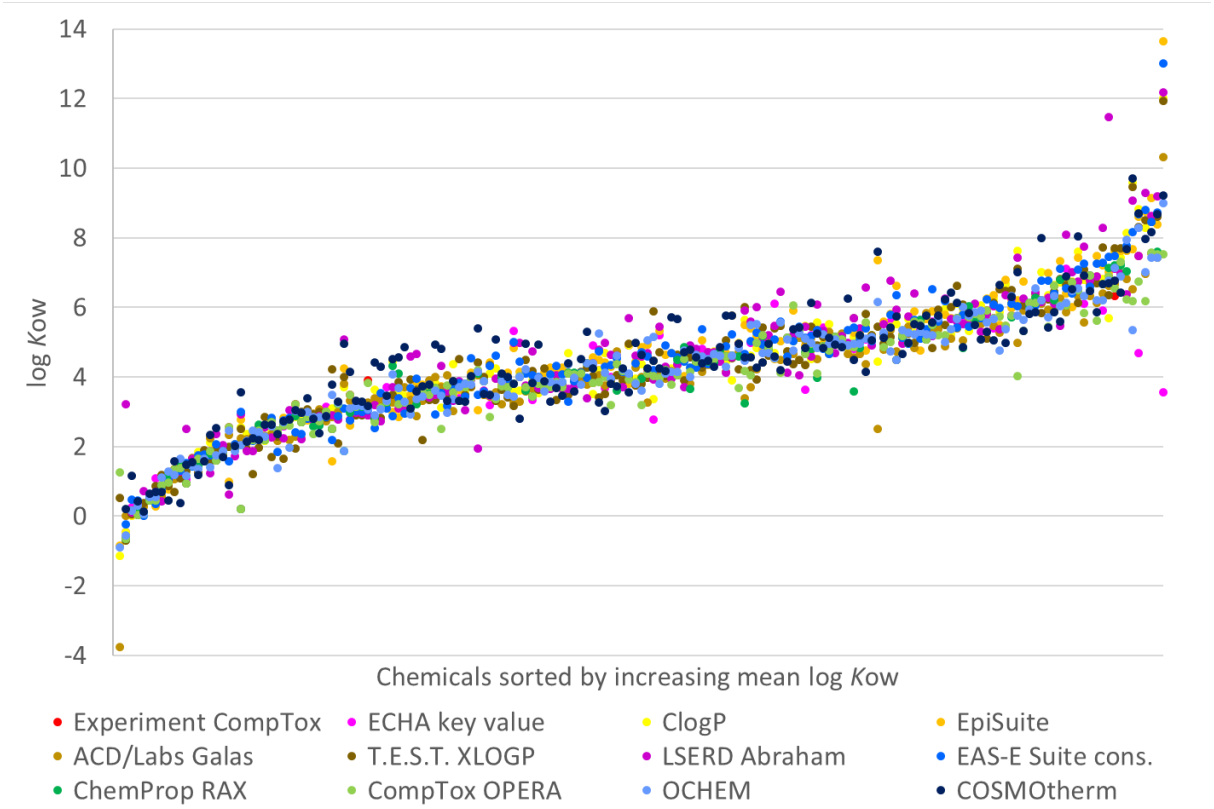


Table 12: **Performance of different methods for determining log K_{OW} values of 173 case study chemicals (availability: number of data points (n) obtained, accuracy: correlation coefficient (r^2), percentage of n within ± 0.3 log units of the overall mean log K_{OW} value (% ± 0.3) and root mean square error (RMSE), variability: largest negative deviation (Δ MIN) underestimating and largest positive deviation (Δ MAX) overestimating log K_{OW}, in relation to the overall mean log K_{OW} value). Adapted from Nendza et al. 2025. For data and details see the supplementary material [log K_{OW}].**

Method	Tool	n	r^2	% ± 0.3	RMSE	Δ MIN	Δ MAX
Experimental data	EpiSuite	113	0.97	75.2	0.34	-1.79	1.00
	CompTox experimental	105	0.97	80.0	0.30	-1.79	0.83
	VEGA experimental	94	0.98	80.9	0.23	-0.50	0.67
	EAS-E Suite experimental	108	0.97	78.7	0.32	-1.79	1.00
	Ulrich et al. experimental	102	0.97	80.4	0.29	-1.79	0.70
	ECHA key value	77	0.74	64.9	1.00	-7.18	1.56
Group contribution methods	ClogP	173	0.96	65.9	0.43	-1.27	1.90
	EpiSuite	173	0.95	56.6	0.54	-1.42	2.91
	VEGA Meylan kowwin	173	0.91	44.5	0.73	-3.59	2.91
	EAS-E Suite kowwin	165	0.96	57.0	0.50	-1.42	2.87
	EAS-E Suite IFQSAR	165	0.93	50.9	0.56	-1.47	1.72

Method	Tool	n	r^2	% ± 0.3	RMSE	Δ MIN	Δ MAX
	T.E.S.T. XLOGP	173	0.94	50.9	0.51	-1.21	1.93
	T.E.S.T. ALOGP	173	0.93	48.0	0.54	-1.38	1.61
	VEGA MLogP	173	0.82	30.1	1.02	-4.28	0.63
	ChemProp Consensus	172	0.94	47.7	0.50	-1.29	1.33
	MolInspiration	173	0.94	70.5	0.42	-2.56	1.93
	DataWarrior	173	0.87	22.5	0.90	-5.34	1.68
	ChemAxon	173	0.94	63.0	0.49	-2.98	1.46
	ACD/Labs Galas v.1.2	173	0.95	71.7	0.43	-2.53	0.78
	ACD/Labs Classic	172	0.93	61.0	0.52	-2.06	2.13
Linear solvation energy relationships (LSER)	ChemProp LSER	167	0.87	48.5	0.77	-3.32	2.72
	UFZ LSERD Abraham Absolv descr.	111	0.97	79.3	0.32	-1.41	1.01
	UFZ LSERD Abraham calc. descr.	171	0.87	53.8	0.78	-1.73	4.50
	UFZ LSERD Goss Absolv descr.	111	0.97	68.5	0.36	-1.28	1.14
	UFZ LSERD Goss calc. descr.	171	0.87	40.4	0.77	-1.98	3.98

Method	Tool	n	r^2	% ± 0.3	RMSE	Δ MIN	Δ MAX
	EAS-E Suite ppLFER	24	0.94	58.3	0.53	-0.58	1.73
Automated read-across (k-nn)	ChemProp Read-across	123	0.95	75.6	0.38	-1.79	1.00
	CompTox OPERA	166	0.91	67.5	0.55	-3.21	2.66
	EAS-E Suite Opera k-nn	25	0.99	96.0	0.15	-0.34	0.24
Deep learning neural networks	OCHEM ALogPS 2.1	173	0.93	64.7	0.48	-2.87	3.07
	OCHEM ALogPS 3.0	173	0.94	68.2	0.43	-2.34	1.24
	DNN complex Ulrich et al.	173	0.90	59.0	0.57	-2.36	3.33
	DNN simple Ulrich et al.	173	0.92	59.5	0.55	-2.89	2.75
Quantum-chemistry	COSMOtherm neutral chemicals at pH 7.4	69	0.94	50.7	0.58	-1.52	2.03
	COSMOtherm all chemicals	172	0.87	45.9	0.70	-1.51	2.56
Consensus	EAS-E Suite Consensus	165	0.97	61.2	0.44	-0.82	2.27
	ACD/Labs Physchem Profiler	173	0.97	80.9	0.30	-2.53	0.54

5.4.3 Consolidation of log K_{OW} values

Consensus modelling, e.g. (Tebes-Stevens et al. 2018; Ahlers et al. 2019; Hodges et al. 2019; Li et al. 2022), reduces uncertainties and is superior to the traditional preference for a single measured log K_{OW}. Combining multiple estimates from different methods in a WoE or averaging approach, takes advantage of their strength while compensating outliers (Nendza et al. 2025). The recommended procedure is comprehensive whenever the entire range of log K_{OW} values is considered. However, if the question is whether a log K_{OW} value is below or above a certain threshold, e.g. 3 or 4.5 as in the case of bioaccumulation assessment, a reduced approach may be sufficient.

For comprehensive log K_{OW} estimation, we recommend using at least 5 log K_{OW} data, including validated experimental data and estimates calculated by different conceptual approaches, such as a group contribution (atom/fragment) method, a linear solvation energy relationship (LSER), automated read-across and a deep learning neural network. If available, a quantum chemical method may be included as well. Estimates from other methods with alternative approaches are a welcome addition. Provided that the computational methods for log K_{OW} estimation fulfil the OECD criteria for scientific validity (OECD 2007) and that the target compounds fit within their AD, reliable consolidated log K_{OW} with reasonable uncertainties can be obtained. The OECD QSAR Assessment Framework is a comprehensive documentation tool for the assessment of the reliability of QSAR predictions in regulatory contexts (OECD 2023; Nendza et al. 2025).

The reduced approach takes advantage of the fact that the margin of error in log K_{OW} values is often within 1 log unit. Therefore, two independent log K_{OW} estimates may suffice if they are both at least one log unit below or above a given threshold value. Only if the available log K_{OW} values are within ± 1 log units of the respective threshold may further data be required to come to a robust conclusion. When well justified, deviations from the systematic approach are both valid and meaningful.

Box 1 outlines the systematic WoE or averaging approach that combines estimates from at least 5 different tools (6-step-procedure). The crucial first step is the verification of the chemical structures and their proper description, mostly by SMILES, with due regard to tautomers and speciation. This is usually more time-consuming than the subsequent input to several freely available/public domain, commercial or in-house software to calculate log K_{OW} for the non-ionised substance. If the model output includes information about reliability of estimates, such as AD coverage and results for similar substances, highly rated estimates are preferred. Erroneous and uncertain estimates, that may be due to AD violation, and obvious outliers should be discarded. Indicators of the quality of estimates, such as reliability ratings and similarity indices, may further be input to weight the evidence. The (weighted) mean of the remaining estimated and reliable experimental data is then the consolidated log K_{OW} (Nendza et al. 2025).

Box 1: Consensus modelling for consolidated log K_{OW} values. Reproduced from Nendza et al. 2025.

6-step-procedure:

(1) Verify chemical identity (2D and 3D structure, not only CAS), with due regard to tautomers and speciation.

(2) Check AD compliance, excluding inorganics, organometallics, PFAS, siloxanes, surfactants. Furthermore, specific AD limitations of individual models must be considered.

(3) Calculate/determine as a minimum 5 log K_{OW} (non-ionised chemical) with different independent methods, including quality-controlled experimental data and results from a group contribution (atom/fragment) method, a linear solvation energy relationship (LSER), automated read-across and a deep learning neural network. Additionally, estimates by other approaches, e.g. quantum-chemistry-based, may be included.

(4) Analyse results for outliers, eliminate obvious misestimates.

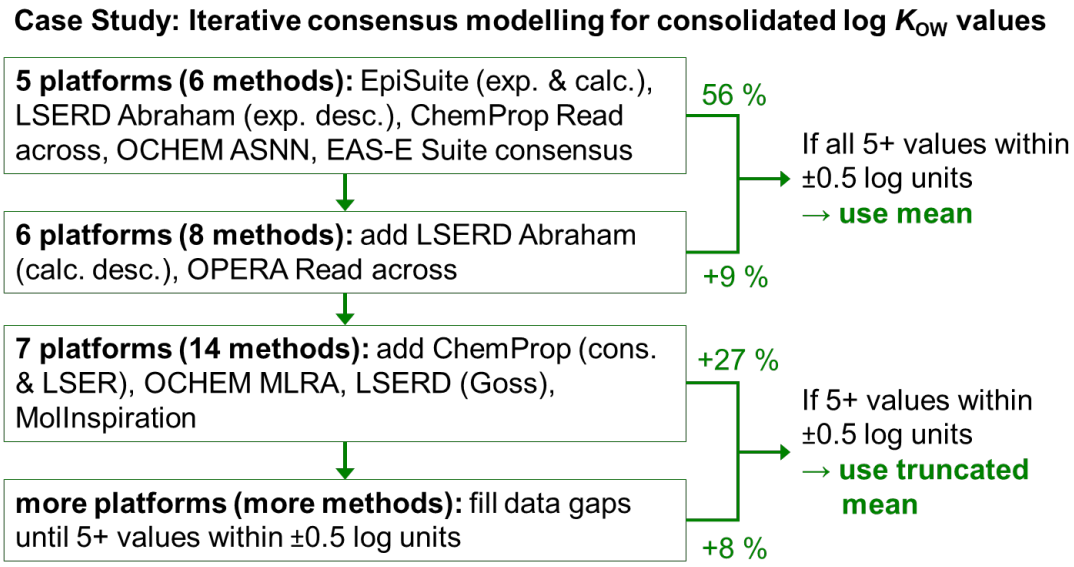
(5) Apply WoE or averaging approach → **consolidated log K_{OW}**.

(6) If needed, use p*K*a for log D adjustment (according to the Henderson-Hasselbalch relationship).

For ionisable substances, the log K_{OW} addressing the neutral form of chemicals, may be misleading. To account for different partition patterns of neutral and ionic species, the actual hydrophobicity of an ionisable compound under given pH conditions may be expressed by the distribution coefficient log D. Based on non-ionised and ionised fractions according to the Henderson-Hasselbalch relationship, p*K*a is used for adjustment (Schüürmann et al. 2007; Köhler et al. 2023).

We demonstrate the performance of iterative consensus modelling for consolidated log K_{OW} values with 173 case study chemicals (Figure 8). Starting with five freely available/public domain tools: EPISuite (US EPA 2012) for an experimental value and a group contribution method, the Abraham Absolv model in UFZ LSERD (Ulrich et al. 2017), ChemProp Read-across (UFZ 2021), OCHEM ALogPS 3.0 (Sushko et al. 2011) and the consensus log K_{OW} of EAS-E Suite (ARC 2023), we get 5+ values within ± 0.5 log units for the majority of the case study chemicals (56%). Their means are the consolidated log K_{OW} values. If less than 5 values were obtained, e.g. because experimental LSER descriptors are lacking or due to limited read-across database, we added the results of the Abraham model using calculated descriptors in UFZ LSERD (Ulrich et al. 2017) and OPERA from CompTox Chemicals Dashboard (US EPA 2023), allowing for consolidated log K_{OW} values based on 5+ values within ± 0.5 log units for another 9% of the case study chemicals. For the remaining 35% of the case study chemicals, the variability of the different log K_{OW} estimates is larger than ± 0.5 log units. Then, we applied more methods available with UFZ LSERD (Ulrich et al. 2017), ChemProp (UFZ 2021), OCHEM ALogPS 3.0 (Sushko et al. 2011) and added results from MolInspiration (2023). Among these, we obtained consolidated log K_{OW} values by calculating the truncated mean, i.e. the mean of all (5+) data within ± 0.5 log units of the median (Steinmetz et al. 2014) for another 27% of the case study chemicals. For the remaining 8%, we obtained more log K_{OW} estimates by more methods to fill data gaps until 5+ values within ± 0.5 log units allowed consensus modelling (Nendza et al. 2025).

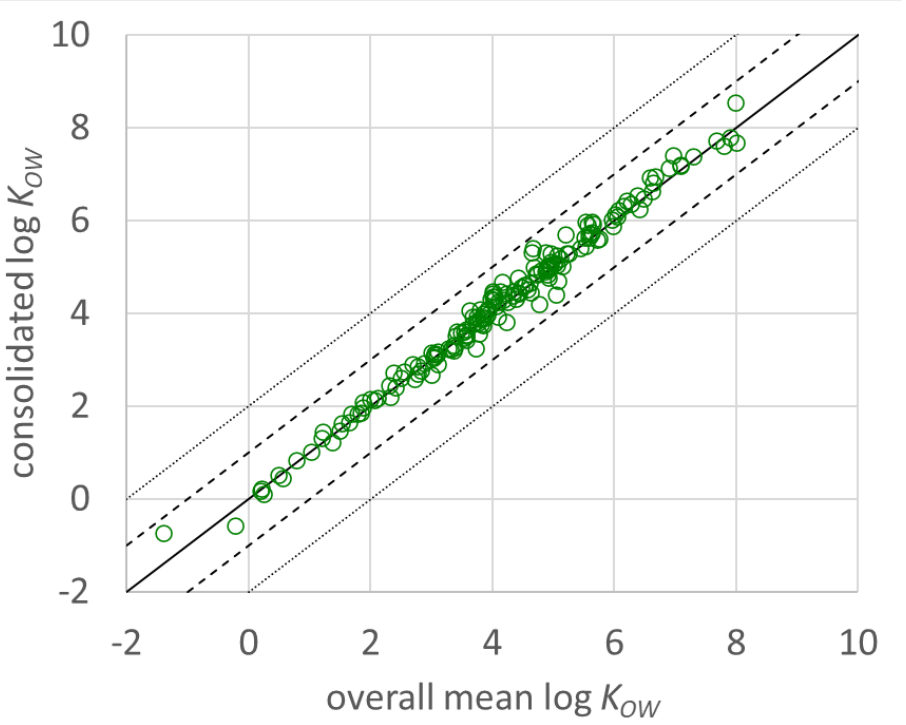
Figure 8: **Iterative consensus modelling for consolidated log K_{OW} values: Procedure and results with 173 case study chemicals. Reproduced from Nendza et al. 2025.**



The consolidated log K_{OW} values based on (truncated) means correlate with the overall mean log K_{OW} values across all methods (r^2 = 0.987, slope = 1.02, intercept = 0.01) (Figure 9). The agreement is close (RMSE = 0.23) and 82% of the case study chemicals deviate by less than 0.3 log units. Only 9 substances deviate by more than 0.5 log units, yet all deviations are less than one log unit (Nendza et al. 2025).

For external validation, we used the 60 reference substances in OECD TG 117 (OECD 2022b) and compared their consolidated log K_{OW} according to Box 1 with the recommended log K_{OW} in the test guideline. Overall, the correlation is excellent (r^2 = 0.98), with deviations < 0.2 log units for more than 80% of the substances. Only 2 chemicals deviate by more than 0.5 log units. One is the weak acid 2,4-dinitro-6-sec-butylphenol (CAS 88-85-7), for which both the consolidated log K_{OW} (3.58) and the experimental value (3.56) in the CompTox Chemicals Dashboard (US EPA 2023) are significantly lower than the reference value (4.1) in OECD TG 117 (OECD 2022b). The other is dodecanoic acid, better known as lauric acid which is a surfactant. Both the consolidated K_{OW} (4.87) and the experimental value (4.6) in the CompTox Chemicals Dashboard (US EPA 2023) are well above the reference value (4.2) in OECD TG 117 (OECD 2022b). This deviation may be explained with surfactants being outside the AD of most experimental and computational methods for the determination of log K_{OW}. Except for these two outliers, this comparative exercise can be regarded as a successful external validation that further supports the accuracy and robustness of the consolidated log K_{OW} Nendza et al. 2025).

Figure 9: **Consolidated log K_{OW} data of 173 case study chemicals compared to the overall mean log K_{OW} across all (experimental and estimation) methods. The solid line is the 1:1 line; the dashed and dotted lines indicate deviations of 1 and 2 log units, respectively. For data details see the supplementary material (log K_{OW}, Table SI-8). Reproduced from Nendza et al. 2025.**



5.4.4 Conclusions

Three key findings emerge from our analyses: (1) measured and calculated log K_{OW} reveal considerable variability up to one log unit and more, without systematic pattern for different chemicals, (2) no method (experimental or computational) is consistently superior (or inferior) for determining log K_{OW}, and (3) the consolidated log K_{OW}, i.e. the (weighted) average of valid measured and estimated data per chemical, is robust and highly reproducible (usually within 0.2 log units), scientifically credible, as well as cost-efficient (Nendza et al. 2025).

A reduced approach can be well justified to assess substances against log K_{OW} thresholds, e.g. 3 and 4.5 in bioaccumulation assessment, and such deviations from the systematic approach may be both valid and meaningful.

More detailed information, the data, statistics and figures are provided in the published paper (https://doi.org/10.1186/s12302-025-01072-2) and in the supplementary material [log K_{OW}].

5.5 Physicochemical parameters for bioaccumulation assessment: log K_{OA}

In the assessment of the bioaccumulation potential of organic chemicals, the 1-octanol/air partition coefficient K_{OA} covers the respiratory elimination in air-breathing organisms (Baskaran and Wania 2023). Modelling results from Gobas et al. (2003) indicate that persistent chemicals with log K_{OA} > 5 and log K_{OW} > 2 have the potential to bioaccumulate in terrestrial food chains. In combination with biotransformation half-life < 70 d, Wania et al. (2022) suggested to use log K_{OA} < 5 and log K_{OW} < 2 to identify organic chemicals not subject to bioaccumulation in air-breathing organisms.

5.5.1 Estimation of log K_{OA} values

Direct measurements of the K_{OA} by generator column or gas chromatography are available for only about 700 substances (Baskaran et al. 2021, Ebert et al. 2023b), about twice as many K_{OA} values were derived from experimental K_{OW} and K_{AW} according to the thermodynamic triangle approach:

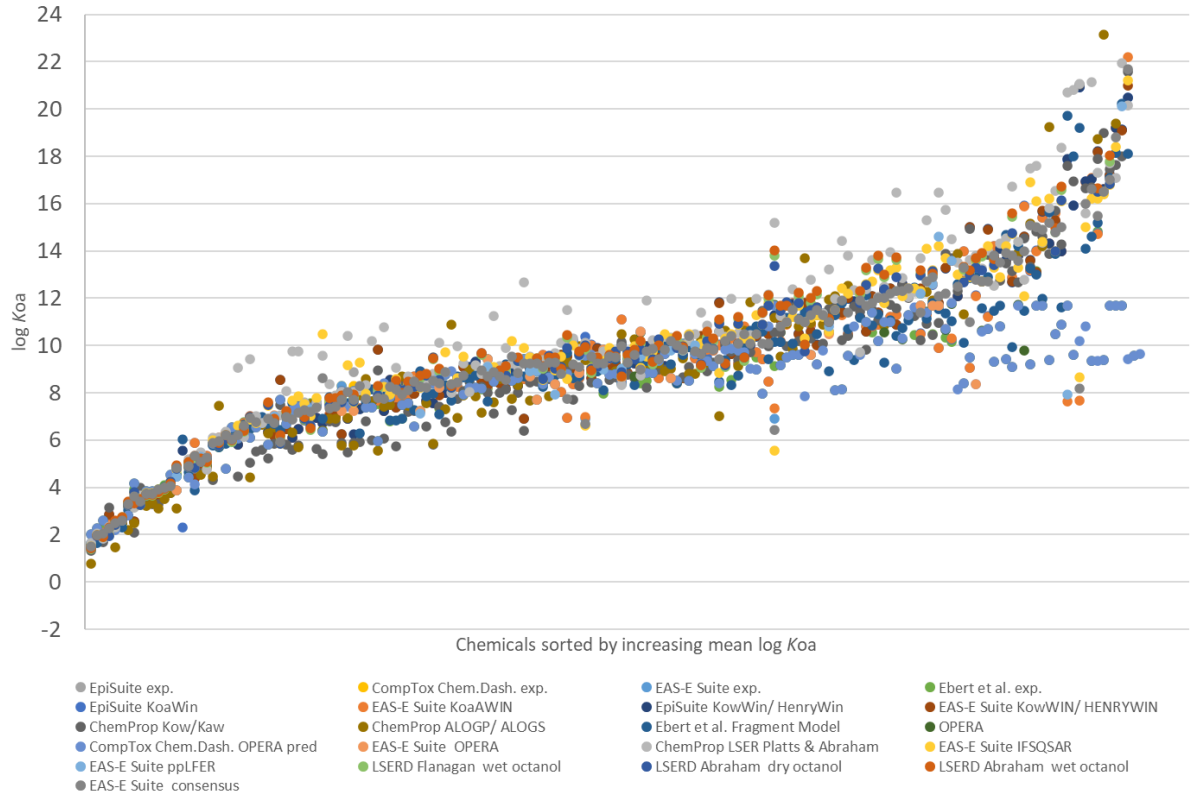
$$K_{OA} \approx \frac{K_{OW}}{K_{AW}} \qquad \text{Eq. 3}$$

Estimated log K_{OA} values also use Eq. 3, but one or both input parameters are calculated instead of measured. This method is included, for example, in EpiSuite (US EPA 2012), ChemProp (UFZ 2021) and EAS-E Suite (ARC 2023). Alternative techniques for log K_{OA} estimation include fragment (group contribution) methods in which the log K_{OA} of a compound is the sum of the incremental contributions of its constituents with constitutional correction factors, e.g. by Ebert et al. (2023). Linear solvation energy relationships (LSER) are implemented, for example, in ChemProp (UFZ 2021), EAS-E Suite (ARC 2023) and UFZ LSERD (Ulrich et al. 2017). An automated read-across model is OPERA (Mansouri et al. 2018), included also in EAS-E Suite (ARC 2023) and the CompTox Chemicals Dashboard (US EPA 2023).

5.5.2 Variability of log K_{OA} values

Because the various approaches used to estimate log K_{OA} differ in their underlying methods, each has distinct strengths and limitations, which can result in varying levels of accuracy for different classes of chemicals—some of which may fall within or outside the applicability domain of a given method. Understanding the accuracy, reliability, and variability of log K_{OA} estimates is particularly important when substances are evaluated against (regulatory) thresholds that rely on this parameter. The systematic comparison of log K_{OA} values, measured and calculated, has revealed that deviations of more than one log unit are the rule rather than the exception (Figure 10). Least scatter is seen with log K_{OA} values below 6. The largest deviations occur at higher log K_{OA} values above 9. Outliers occur with all methods and tools, regardless of the substance class. No method is throughout superior (or inferior) for estimating the log K_{OA} of the diverse case study chemicals (Table 13). To decide whether a log K_{OA} value is correct or an outlier, consolidation by further independent measurements and/or estimates is recommended.

Figure 10: **Variability of log K_{OA} estimates for 173 case study chemicals obtained by different experimental and computational methods. The x-axis represents the 173 case study chemicals sorted by increasing mean log K_{OA}, the y-axis reveals the range of the log K_{OA} data for each chemical. See Table 13 for details on the performance of the tools and references.**



Source: own illustration, AL-Luhnstedt

Table 13: Performance of different methods for determining log K_{OA} values of 173 case study chemicals (availability: number of data points (n) obtained, accuracy: percentage of n within ± 0.5 log units of the overall mean log K_{OA} value (% ± 0.5), variability: largest negative deviation (Δ MIN) underestimating and largest positive deviation (Δ MAX) overestimating log K_{OA} in relation to the mean log K_{OA} value).

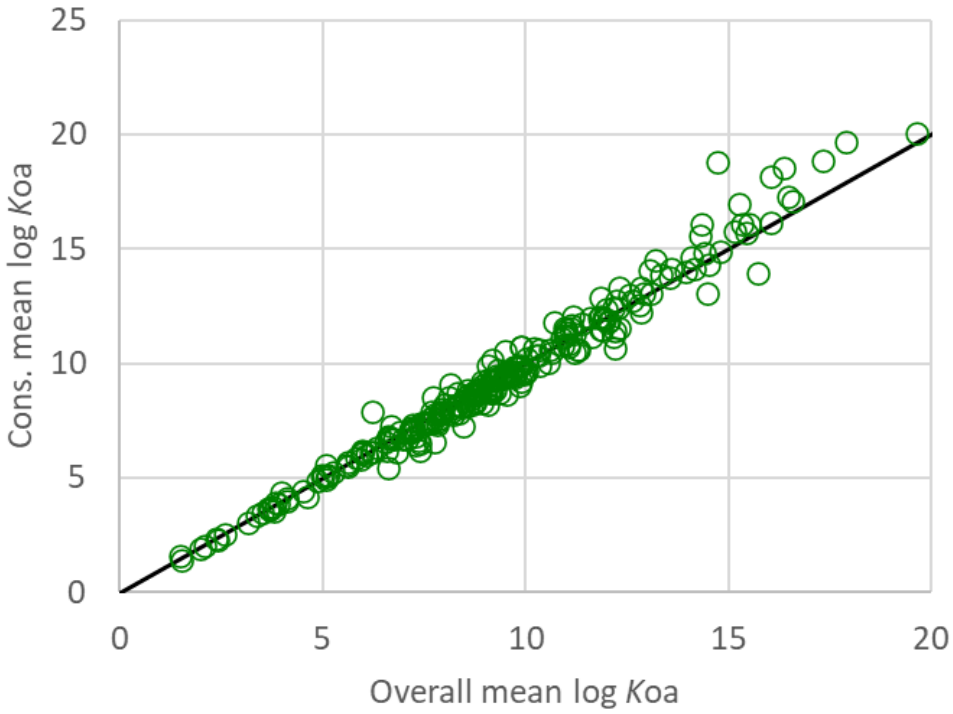
Method	Tool	n	% ± 0.5	Δ MIN	Δ MAX
Experimental data	EPISuite (US EPA 2012)	21	90.5	-0.31	1.29
	CompTox Chemicals Dashboard (US EPA 2023)	19	84.2	-0.31	0.78
	EAS-E Suite (ARC 2023)	19	84.2	-3.71	0.78
	Ebert et al. (2023)	84	66.7	-2.05	0.77
Thermodynamic triangle approach (K_{OA} = K_{OW}/K_{AW})	EPISuite (US EPA 2012) KoaWin	172	58.7	-3.25	7.17
	EAS-E Suite (ARC 2023) KoaWin	161	59.6	-7.65	7.18
	EPISuite (US EPA 2012) KowWin/HenryWin	172	59.9	-1.94	7.18
	EAS-E Suite (ARC 2023) KowWin/HenryWin	115	64.3	-1.94	2.68
	ChemProp (UFZ 2021) Kow/Kaw	172	40.1	-2.47	>10
	ChemProp (UFZ 2021) ALOGP/ALOGS	114	43.9	-2.86	7.08
Group contrib. methods	Ebert et al. (2023)	173	60.1	-3.69	4.97
	EAS-E Suite (ARC 2023) IFQSAR	162	58.6	-6.65	>10
Linear solvation energy relationships (LSER)	ChemProp (UFZ 2021)	172	42.4	-1.65	>10
	EAS-E Suite (ARC 2023) ppLFER	28	75.0	-6.81	2.42
	UFZ LSERD (Ulrich et al. 2017) Flanagan wet octanol	101	54.5	-0.92	3.22
	UFZ LSERD (Ulrich et al. 2017) Abraham dry octanol	101	64.4	-1.19	2.75
	UFZ LSERD (Ulrich et al. 2017) Abraham wet octanol	101	49.5	-0.91	3.43
Automated read-across (k-nn)	OPERA (Mansouri et al. 2018)	143	46.2	<-10	1.11
	CompTox Dashboard (US EPA 2023) OPERA	170	47.1	<-10	1.13
	EAS-E Suite (ARC 2023) OPERA	25	52.0	-4.17	1.13
Consensus	EAS-E Suite (ARC 2023) consensus	161	77.6	-7.15	6.68

5.5.3 Consolidation of log K_{OA} values

The major variability of measured and calculated log K_{OA} values by more than one log unit calls for consensus modelling to reduce uncertainties and consolidate variable, possibly conflicting, results. The recommended procedure is comprehensive whenever the entire range of log K_{OA} values is considered. However, if the question is whether a log K_{OA} value is below or above the threshold of 5, as in the case of bioaccumulation assessment, a reduced approach is sufficient.

The reduced approach takes advantage of the fact that the margin of error in log K_{OA} values below 7 appears limited to 1 log unit (Figure 11). Thus, if the available log K_{OA} values are all within 1 log unit and less than 4, it is sufficiently likely that their (weighted) mean is below the threshold of 5 to identify organic chemicals with low potential for bioaccumulation in air-breathing organisms. Estimates between 4 and 6 may require more data to come to a robust conclusion. If a precise log K_{OA} above 5 is not needed, the calculations end here.

Figure 11: Consolidated mean log K_{OA} of the case study chemicals compared to the overall mean log K_{OA} across all methods. The solid line is the 1:1 line. For data and details see the supplementary material [log K_{OA}; Table SI-7].



Source: own illustration, AL-Luhnstedt

The tiered procedure comprises 5 to 7 steps and combines estimates from at least 2 different methods:

(1) The starting point is always to verify the chemical identity (2D and 3D structure, not only CAS), with due regard to tautomers and speciation.

(2) Suitable models from freely available / public domain, commercial or in-house software to calculate log K_{OA} are identified and AD compliance is checked, considering specific AD limitations of individual models.

(3) 2 log K_{OA} values (non-ionised chemical) are calculated/determined with different independent methods, including, e.g., quality-controlled experimental data and results from the

thermodynamic triangle approach ($K_{OA} = K_{OW}/ K_{AW}$), group contribution (atom/fragment) methods, linear solvation energy relationships (LSER), automated read-across (k-nn), quantum-chemistry, or deep learning neural networks.

(4) If the model output includes information about reliability of estimates, such as AD coverage, confidence intervals and results for similar substances, highly rated estimates are preferred. Uncertain estimates, that may be due to AD violation, and obvious outliers should be excluded.

(5) If 2 reliable log K_{OA} values are within 1 log unit and less than 4, their mean can be used with sufficient confidence → **consolidated log K_{OA}**.

(6) If the 2 log K_{OA} values are between 4 and 6, more log K_{OA} values using different independent methods (steps (3) and (4)) may be needed until a robust conclusion is feasible → **consolidated log K_{OA}**.

(7) If the 2 log K_{OA} values are above 5, the calculations end here if a precise log K_{OA} above 5 is not needed. Else, more log K_{OA} values are determined using different independent methods (steps (3) and (4)) until at least 4 reliable values within 1 log unit are obtained. Their mean can be used with sufficient confidence → **consolidated log K_{OA}**.

5.5.4 Conclusions

Three key findings emerge from the analyses: (1) measured and calculated log K_{OA} reveal considerable variability up to one log unit and more, (2) no method (experimental or computational) is consistently superior for determining log K_{OA}, and (3) the consolidated log K_{OA}, i.e. the (weighted) average of all valid measured and estimated data per chemical, is robust and highly reproducible, scientifically credible, as well as cost efficient.

The limited variability of low log K_{OA} (within one order of magnitude) allows a reduced approach to assess substances against the threshold of 5 for unlikely bioaccumulation in air-breathing organisms.

More detailed information, the data, statistics and figures are provided in Annex A.11 and in the supplementary material.

5.6 Physicochemical parameters for bioaccumulation assessment: p*K*a

5.6.1 Acid dissociation constants: Data availability, estimation methods and uncertainties of p*K*a

The extent to which an organic acid or base bioaccumulates in aquatic organisms depends primarily on two physicochemical properties, the 1-octanol/water partition coefficient log K_{OW} and the acid dissociation constant p*K*a (Schwarzenbach et al. 1993; Nendza 1998; Nendza and Hermens 1995; Schüürmann et al. 2007; Köhler et al. 2023). Under environmental conditions, i.e. water bodies having pH between 5 and 9, only weak to moderate acids (p*K*a > 4) and bases (p*K*a(H) < 10) can be partially ionised with pH-dependent partitioning profiles (Table 14). The bioaccumulation of strong acids (p*K*a < 4) and bases (p*K*a(H) > 10) is not influenced by the ambient pH, as it is only due to the ionised species of the compounds. Strong acids and bases are therefore not considered here.

Table 14: Distribution of ionised (> 10%) and unionised species of organic acids and bases under environmental conditions.

	Ambient pH		
	< 5	**5 - 9**	**> 9**
Neutral substances	unionised	unionised	unionised
Strong acid, p*K*a < 4	(partly) ionised (-)	ionised (-)	ionised (-)
Acid, p*K*a 4 - 10	(partly) ionised (-)	(partly) ionised (-)	(partly) ionised (-)
Weak acid, p*K*a > 10	unionised	unionised	(partly) ionised (-)
Weak base, p*K*a(H) < 4	(partly) ionised (+)	unionised	unionised
Base, p*K*a(H) 4 - 10	(partly) ionised (+)	(partly) ionised (+)	(partly) ionised (+)
Strong base, p*K*a(H) > 10	ionised (+)	ionised (+)	(partly) ionised (+)

Experimental methods for determining p*K*a include titration, conductometry, electrometry, photometry and magnetic resonance. Compilations of p*K*a values are available, e.g. (Perrin et al. 1981; Slater et al. 2014; OECD 2023; ECHA 2023). The p*K*a values reported in the literature for a given organic acid or base may vary by 0.3 p*K*a units, depending on the type of measurement and the conditions chosen (Schwarzenbach et al. 1993). Practical examples reveal even larger ranges, e.g., in the ECHA registration dossier of ethylparaben (CAS 120-47-8), the key value of 8.4 was obtained by considering several reliable values in the range of 8.18 to 8.9 in a weight-of-evidence approach.[1]

Computational methods for estimating p*K*a values are often based on linear free energy relationships (LFER), using Hammet-type equations for chemical classes sharing the same ionizable functional groups, such as phenols, benzoic acids, anilines, substituted pyridines, etc. (Perrin et al. 1981; Schwarzenbach et al. 1993; Nendza 1998; Nendza and Hermens 1995; Lyman et al. 1982; ACD/Labs 2023). An alternative is automated read-across, e.g. (Mansouri et al. 2018; US EPA 2023), using a number of close analogues (k-nearest neighbours (k-nn)) with a minimum chemical similarity distance. Further options for p*K*a calculations are SPARC (ARChem 2010), using computational algorithms based on fundamental chemical structure theory, and Chemicalize (Chemaxon 2024), available for small molecules (≤ 12 non-hydrogen atoms) with ChemSpider (Royal Society of Chemistry 2018). Machine learning methods for p*K*a prediction have been described by, e.g., (Mansouri et al. 2019; Navo & Jiménez-Osés 2021; Luo et al. 2024). While there are quite a number of proprietary software packages for the prediction of p*K*a, the availability of free and open-source programmes for this purpose is limited.

5.6.2 Case study (p*K*a)

A case study was carried out as part of this study having three main objectives: (1) to obtain several p*K*a values by different methods for the case study chemicals to exemplify the variability of p*K*a values, (2) to analyse the variability of p*K*a estimates regarding the underlying methods and the different chemical classes, and (3) to recommend approaches to obtain reliable and robust p*K*a estimates for the hazard and risk assessment of chemicals.

[1] https://echa.europa.eu/registration-dossier/-/registered-dossier/13843/4/22 (accessed 09/04/2024).

5.6.2.1 Chemical dataset

Among the 231 chemicals selected for this case study are 25 strong acids (p*K*a < 4) and 9 strong bases (p*K*a(H) > 10). 33 acids and 26 bases have p*K*a or p*K*a(H) between 4 and 10, respectively. For data and details see the supplementary material [pKa; Tables SI-1, SI-2, SI-3].

5.6.2.2 Data quality of experimental and calculated p*K*a values

Major issues in data quality of experimental and calculated p*K*a values are related to incorrect documentation, e.g. the reporting of *K*a values instead of p*K*a values, or the mistaking of p*K*a (acidic) for p*K*a (basic), and vice versa. The former is found repeatedly in the ECHA database (ECHA 2023), while the latter is particularly seen with the ACD/Labs estimates in the CompTox Chemicals Dashboard (US EPA 2023). In EAS-E Suite (ARC 2023), it is noticeable that substances without stored p*K*a values are incorrectly specified as 'neutral at pH 7.4'. Examples of this are isophthalic acid (CAS 121-91-5) and nonylamine (CAS 112-20-9).

Overall, our experience from this case study is that the identification of reliable p*K*a values requires expertise so that incorrect or misleading data can be excluded.

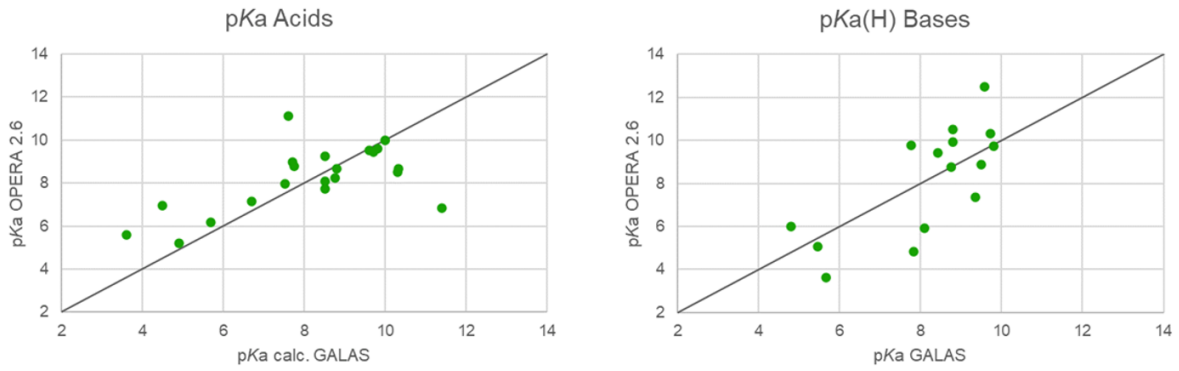
5.6.2.3 Variability of experimental and calculated p*K*a values

Curation of the available data for the 33 acids and 26 bases among the case study chemicals with p*K*a or p*K*a(H) between 4 and 10, results in average ranges of 1.5 log units. For about half of the acids (17 (52%)) and bases (13 (50%)), these p*K*a values are within 1 log unit, while for about one third of the acids (9 (27%)) and bases (9 (35%)) they vary by more than 2 log units.

5.6.2.4 Evaluation of the variability of p*K*a estimates with regard to the underlying methods and in terms of different chemical classes

The performance of LFER (Hammet equation)-based methods using ACD/Labs GALAS (ACD/Labs 2023) and automated read-across (k-nn) using OPERA 2.6 from the CompTox Chemicals Dashboard (US EPA 2023) were compared (Figure 12). The agreement between the estimates is reasonable for the acids (with 4 major outliers), while the estimates for the bases scatter without systematic pattern. Within well-defined chemical classes such as phenols or alkylamines among the case study chemicals, consistent p*K*a estimates are obtained. For other substance classes, however, such analyses are not possible due to the limited data set.

Figure 12: Comparison of p*K*a estimates for acids (left panel) and p*K*a(H) estimates for bases (right panel) by the LFER (Hammet equation)-based method using ACD/Labs GALAS (ACD/Labs 2023) and automated read-across (k-nn) using OPERA 2.6 from the CompTox Chemicals Dashboard (US EPA 2023). The solid line is the 1:1 line.



Source: own illustration, AL-Luhnstedt

5.6.2.5 Recommendations for obtaining reliable and robust p*K*a estimates for the hazard and risk assessment of chemicals

If reliable measured p*K*a values are not available, consistent calculated data by two or more independent methods, e.g. by the LFER (Hammet equation)-based method using ACD/Labs GALAS (ACD/Labs 2023) and automated read-across (k-nn) using OPERA 2.6 from the CompTox Chemicals Dashboard (US EPA 2023), are the recommended alternative. Provided that the computational methods for p*K*a estimation fulfil the OECD criteria for scientific validity (OECD 2007) and that the target compounds fit within their AD, the OECD QSAR Assessment Framework can be used to document reliable QSAR predictions for regulatory applications (OECD 2023a).

If no acceptable p*K*a values can be obtained in this way, there is still the option of a worst-case approach to the bioaccumulation assessment of the non-dissociated substance. Further data requirements may arise only if thresholds are exceeded.

More detailed information, the data, statistics and figures are provided in chapter A.12 and in the supplementary material [p*K*a; Tables SI-1, SI-2, SI-3].

6 Evaluation of the integrated test- and assessment strategy

The test methods were evaluated regarding their suitability for use in the regulatory frameworks. For comparison of the test methods, 26 substances (Table 3) were selected for which a nearly complete set of data is available including physicochemical data, QSAR BCFs, *H. azteca* and in vitro BCF values as well as fish BCF data. For some substances, also rat in vitro data was available, or new substrate depletion assays were conducted. The bioaccumulation potential of the substances was assessed based on the proposed integrated assessment scheme (see chapter 4). The outcome of the tests was evaluated with regard to their comparability. This did not include an evaluation of all available data, thus the regulatory decisions taken with regard to, for example SVHC or POP identification, may not be reflected here.

6.1 Bioaccumulation assessment of 26 case study chemicals

An overview of the data set of the 26 case study chemicals including information on physicochemical properties, QSAR BCFs, invertebrate and in vitro BCFs as well as in vivo BCF data is shown in Table 15. This dataset is used to evaluate how the results from the fish in vitro tests and the *Hyalella azteca* BCF tests compare with the results from the fish in vivo BCF tests, and how these results could be used in a regulatory context, using the proposed assessment schemes (Scheme 1 & 2). The detailed results for the different substance groups can be found in the breakout tables (Table 16 – Table 22).

For each of the 26 case study chemicals, the physicochemical data (tier 1) and BCF estimates for tiers 2-4 are collected as described in the previous chapters:

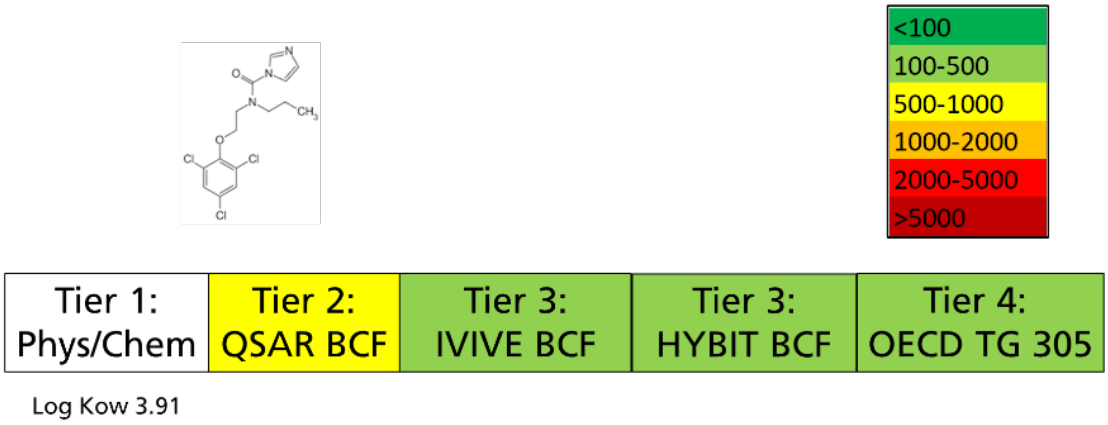
The physicochemical parameters log K_{OW} and log K_{OA} that are used on tier 1 are the consolidated values that were derived in chapters 5.4 and 5.5 and which are listed in supplementary material "log K_{OA}" and "log K_{OW}". Fish QSAR BCFs that are presented in tier 2 of the water breathing assessment tree (Scheme 1) are the mean values of the values calculated in chapter 5.3 (also *cf.* supplementary material "BCF QSAR"), using those QSARs that were within the applicability domain of the respective substance. Fish IVIVE BCFs (tier 3, water breathing organisms in Scheme 1) were calculated as described in chapter A.4.4. If multiple studies were available for one substance, the newest studies or those with the lowest reported starting concentration were selected (*cf.* supplementary material "Substance data endpoints"). HYBIT BCFs (tier 3, water breathing organisms in Scheme 1) are kinetic BCFs normalised to 3% lipid content, if not stated differently in the follow-up tables. If multiple HYBIT BCFs were available, studies were omitted from the comparison if no lipid values were reported or the *H. azteca* population in the test was not only male (*cf.* supplementary material "Substance data endpoints"). If one study reported multiple BCFs, the median BCF value is reported. For in vivo fish BCFs (tier 4, water breathing organisms in Scheme 1) the full range of BCFs is reported. The IVIVE BMFs based on in vitro hepatic biotransformation rates in rats were calculated as described in chapter A.5.6.

To facilitate the readability and evaluation of the data, colour codes are used as indicated in the legend of the Table 15. The aim of the project is not to identify substances as B or non-B – this has been done elsewhere for these substances – but to evaluate if a decision on the bioaccumulation of a given substance can be taken safely on the basis of NAMs, without false negatives, or in other words, without lowering the level of protection. This also includes a reflection on which test is more appropriate for different substance groups.

This is demonstrated with the example of prochloraz. The results of the different tests are shown in Figure 13. At a first glance, a coherent result across all tiers was obtained for the

moderately hydrophobic compound (log K_{OW} 3.91), with all BCF estimates being below the value of BCF 1000. Nonetheless, no assessment should be made without examining the test results and the respective publication or test report in detail:

Figure 13: B-assessment of prochloraz (Scheme 1).



Source: own illustration, Fraunhofer IME

Tier 1 is defined by the physicochemical parameters. Since prochloraz is a neutral organic substance, the log K_{OW} is the appropriate parameter to look at. Based on the consensus log K_{OW} value of 3.91, a B-assessment is triggered in some regulations. The QSAR BCFs that were calculated range from log BCF 1.91 (BCF = 81) to log BCF 4.12 (BCF = 13234), which results in a mean QSAR BCF of 696 (log BCF = 2.84) (*cf.* SI "BCF_QSAR"). Consequently, the box for tier 2 "QSAR" received the colour yellow. Here, the range of calculated QSAR BCFs is very large, care must be taken in interpreting the result. And additional data is necessary to refine the picture. The SI "Substance data endpoints", chapter "Prochloraz" lists one publication that reports in vitro substrate depletion data for prochloraz, determined using rainbow trout hepatocytes. IVIVE BCFs of 366 and 1140 are reported. Re-calculating the reported in vitro clearance rate using the B-compass fish model for comparison, the IVIVE BCF results in 399 (*cf.* chapter A.8). The paper states that the obtained clearance rate for prochloraz did not follow first order depletion, which is a prerequisite, as a first order depletion is an underlying assumption in the extrapolation models for an IVIVE BCF. Therefore, this IVIVE BCF must be interpreted with caution. Yet, for comparison purposes and due to lack of other information, the IVIVE BCF is used for the comparison and the box tier 3, IVIVE is coloured in light green. The literature screening found 3 sources for HYBIT BCF values for prochloraz, as shown in SI "Substance data endpoints", chapter "Prochloraz". The ring test data of OECD TG 321 is separated into two columns, one that summarises the data for experiments conducted with ^{14}C-radioalbelled prochloraz and one column with non-radiolabelled prochloraz. The first column of the table summarises the data of a publication. If the parent compound is of interest, data based on TRR (total radioactive residue) are not helpful. Therefore, only the data in the first two columns of the table was considered. With a BCF of 185 ± 59 (BCF_{kL}, 3% lipid, ± error) and 119 ± 37 (BCF_{kL}, 3% lipid, mean ± SD), the two HYBIT BCF values fall into the same category and thus the tier 3 box for the HYBIT is graded as light green in Figure 13**Fehler! Verweisquelle konnte nicht gefunden werden.**. In vivo fish data is scarce. The OECD QSAR Toolbox does not contain any BCF data for this substance. 2 BCFs are reported in the Conclusion on Pesticide Peer Review for Prochloraz by the EFSA (EFSA 2011). The first value is a BCF of 371 which is based on ^{14}C-TRR and was measured in bluegill sunfish. The second BCF reported is 197 is also based on ^{14}C-TRR but was measured in rainbow trout. No additional information for these BCF values is available.

Based this information, the box tier 4 is coloured in light green. The database for prochloraz shows that the studies that provide the endpoints can have flaws and not all relevant details are reported. In comparison, the available data is consistent, but nonetheless it is important to keep in mind that blind spots exist. Expert judgement is necessary to evaluate the significance of the drawbacks of studies and the resulting impact on the substance assessment.

Table 15: Data set of the 26 case study chemicals.

		Water breathing organisms						Air breathing organisms
		Tier 1		Tier 2	Tier 3		Tier 4	
Substance name	CAS	Consolidated log Kow	Consolidated log Koa	Mean fish QSAR BCF	IVIVE BCF	HYBIT BCF	Experimental fish BCF range	IVIVE BMF (rat)
Simazine	122-34-9	2.12	8.94	10	7	9	1 - 214	
Azoxystrobin	131860-33-8	3.00	13.9	58	34	5	N/A	
Terbutryn	886-50-0	3.41	9.83	33	50	46	9 - 44	
17a-Ethinylestradiol	57-63-6	3.84	13.6	249	324	47	269 - 331	
β-Hexachlorocyclohexane	319-85-7	3.87	7.59	814	367	887	293 - 1460	24
Prochloraz	67747-09-5	3.91	12.2	696	399	185	197 – 393	
Fluorene	86-73-7	4.06	6.69	781	374 (HEP) 433 (S9)	285	58 - 1288	
Anthracene	120-12-7	4.41	7.44	1252	647 (HEP) 698 (S9)	1800	191 - 6000	
Phenanthrene	85-01-8	4.43	7.45	1436	662 (S9) 799 (HEP)	465	708 - 6118	
Chlorpyrifos	2921-88-2	4.89	9.47	934	509	1163	40 - 3162	0.08
Pyrene	129-00-0	4.91	8.75	1618	144	3050	50 - 1351	0.773
Triclosan	3380-34-5	4.91	11.08	1553	146	N/A	1 - 1995	
Octamethylcyclo-tetrasiloxane (D4)	556-67-2	4.98	4.13	1679	1935	test failed	5012 - 14900	test failed
Trifloxystrobin	141517-21-7	5.08	11.2	579	672	568	372 - 537	
Pentachlorobenzene	608-93-5	5.11	6.53	6603	5499	1913	3981 - 22909	
Methoxychlor	72-43-5	5.25	10.6	2623	1928 (S9) 3404 – 3765 (HEP)	9396	621 - 8300	1.5
o-terphenyl	84-15-1	5.58	8.80	3555	test failed	6086	501 - 12882	
Benzo[a]pyrene	50-32-8	6.07	11.0	3669	787	3911	3 - 631	
PCB 153	35065-27-1	7.08	10.0	195265	13596	155373	8913 - 740000	
UV-234	70321-86-7	7.68	15.32	5761	7464	1670	1286	4 - 33
Ionisable substances								
Fluoxetine	54910-89-3	4.33	10.0	323	0 - 1040	19 - 348	13 - 330	
Diclofenac	15307-86-5	4.35	12.9	376	0 - 944	3 - 73	5 - 9	0 – 0.12
Pentachlorophenol	87-86-5	4.81	9.31	835	14 - 2976	132	195 - 479	
Substances not applicable to schemes 1 & 2								
GenX	62037-80-3	3.26	5.09	5.4	91	7	1 - 8	
PFOS	1763-23-1	4.42	5.81	91.3	N/A	302	255 - 5400	
Lauric acid	143-07-7	4.84	8.51	238	N/A	6	255	

BCF Classification
BCF <100
BCF 100-500
BCF 500-1000
BCF 1000-2000
BCF 2000 - 5000
BCF >5000

6.2 Assessment of bioaccumulation in water-breathing organisms (Scheme 1)

6.2.1 Substances with log K_{OW} < 4.5

Generally speaking, a bioaccumulation assessment is not normally conducted for substances with a "low" log K_{OW}. As described above, which log K_{OW} has been set as a trigger value for the assessment of bioaccumulation, both for hazard assessment and secondary poisoning, differs between the substance regulations. Here we group the substances with a log K_{OW} <4.5, without reference to any specific regulation. The results of the different test systems are compared.

As shown in Table 16, for the lower log K_{OW} substances, e.g. simazine, azoxystrobin, terbutryn and 17α-ethinylestradiol, the resulting BCF values from tiers 2, 3 and 4 match well, and remain well below the trigger values for bioaccumulative or very bioaccumulative substances. With regard to risk assessment and secondary poisoning, also other data would be needed which is out of scope of this study. For β-hexachlorocyclohexane (β-HCH), all BCF data underestimate the true bioaccumulation: this is one of the substances that accumulate much higher in air-breathing organisms than in aquatic organisms, see chapter 6.3, which is why it was identified as POP substance and included in the Stockholm Convention (UNEP 2007). For fluorene very similar results compared to β-HCH were obtained with the different test systems for aquatic bioaccumulation assessment, however, no information on bioaccumulation in air-breathing mammals is available.

The consolidated log K_{OW} for both anthracene and phenanthrene is just below 4.5, ECHA has used a value just above 4.5 for their evaluation. Nevertheless, anthracene was identified by ECHA as bioaccumulative (B) due to the results of the in vivo fish BCF tests carried out with different fish species (BCF 191 – 6000) (ECHA 2008). If only rainbow trout was included, the BCF range would span from 191 – 724. The pre-guideline BCF study with *Hyalella azteca* is from 1983 (Landrum & Scavia 1983). The reported BCF (BCF = 1800) is a steady-state BCF, although steady-state was not reached according to the authors. The BCF is reported as total radioactive residue (TRR), corrected for metabolites. No lipid normalisation was done. The value is above the BCF of 1200 suggested by the ECHA R.11 PBT guidance (ECHA 2023a), thus it could be considered indicative of a need for further analysis, and the study could be evaluated as inconclusive with regard to PBT assessment. However, a correctly estimated lipid normalised steady-state BCF might be higher (e.g. > BCF 2000) and thus lead to another conclusion, which once again emphasises that attention must be paid to the framework conditions of the individual studies. The BCF estimated from the fish in vitro study is in the range of the in vivo fish BCF studies with rainbow trout but does not reflect the high BCF values estimated from studies conducted using other fish species. Phenanthrene was identified as very bioaccumulative (vB) by ECHA due to bioaccumulation in fish and aquatic invertebrates, with very high BCF values in other invertebrates but not in *Hyalella azteca* (ECHA 2018). The BCF study with *Hyalella azteca* (Lee et al. 2002) is a non-guideline test, no lipid normalisation was done. The BCF is given as TRR, the presented BCF is a median value of four values in the range of 440 – 504. The in vitro BCF values both for S9 and hepatocytes are in the same range. Fish BCF values obtained from studies with different fish species ranged from 708-6118. This case shows again that it is necessary to consider all available data in a WoE approach and that substances with a log K_{OW} <4.5 can still have a high bioaccumulation potential.

Table 16: Lipid partitioning (Scheme 1): log K_{OW} < 4.5.

	Tier 1		Tier 2	Tier 3		Tier 4
Chemical	Consolidated log K_{OW}	Consolidated log K_{OA}	Mean fish QSAR BCF	IVIVE BCF	HYBIT BCF	Experimental fish BCF range
Simazine	2.12	8.94	10	7**	9*	1-214
Azoxystrobin	3.00	13.9	58	34	5	N/A
Terbutryn	3.41	9.83	33	50	46	9 - 44
17α-Ethinylestradiol	3.84	13.6	249	324§	47§§	269 - 331
β-Hexachloro-cyclohexane	3.87	7.59	814	367	887	293 - 1460
Prochloraz	3.91	12.2	696	399†	185	197 – 393††
Fluorene	4.06	6.69	781	374 (HEP) 433 (S9)	285ϕ	58 - 1288
Anthracene	4.41	7.44	1252	647 (HEP) 698 (S9)	1800ψ	191 - 6000
Phenanthrene	4.43	7.45	1436	662 (S9) 799 (HEP)	465ϕ	708 - 6118

* TRR-based value

** No biotransformation determinable

§§ Median BCF value, steady-state BCF, no lipid normalisation

† No first order depletion rate determined

†† TRR-based value

ϕ TRR-based value, no lipid normalisation, median value, non –guideline test

ψ Steady-state BCF, steady-state not reached; TRR-data corrected for metabolites; no lipid normalisation

6.2.2 Substances with log K_{OW} 4.5 – 6

Chemicals with a moderate to high hydrophobicity (log K_{OW} 4.5-6) are expected to be in the applicability domain of all test systems considered in Tier 2-4. Predictive QSARs were applied to estimate fish BCFs, based on different methodological approaches (tier 2). A mean BCF fish QSAR value was calculated for each substance. None of the substances (Table 17) had a mean fish QSAR BCF value below 500, which means that, for these substances, there is no clear indication of a low bioaccumulation potential as defined in the previous chapter. Increasing log BCF fish QSAR data with increasing log K_{OW} between 1 and 6 is clearly evident for the "26 substances data set" (chapter 5.3.3.5). However, QSAR predictions for substances of similar hydrophobicity can vary considerably. Mean fish QSAR BCFs of trifloxystrobin (log K_{OW} 5.08) and pentachlorobenzene (log K_{OW} 5.11) were 579 and 6603, respectively. Fish BCF QSAR for both substances were comparable with the results of all other test systems regarding classification of the BCF values.

Mean fish QSAR BCFs were similar to the results of the in vivo fish tests for triclosan and pyrene. In other cases, mean fish QSAR BCFs clearly underestimated the result of the in vivo fish tests as observed for chlorpyrifos and Octamethylcyclotetrasiloxane (D4). which highlights the limited significance of fish QSAR BCFs for substances with medium to high hydrophobicity for regulatory use.

In one case (trifloxystrobin) the results of the in vitro tests confirm the HYBIT BCFs. For octamethylcyclotetrasiloxane (D4) an IVIVE BCF of 1935 was observed. Only 1 out of 3 runs with D4 resulted in linear regressions with an R^2 of > 0.85 which could be due to slow depletion of the test substance. However, the results of fish BCF tests provide clear indications for the substance's bioaccumulation potential with BCF values > 5000. In two cases (methoxychlor, o-terphenyl), HYBIT BCFs > 5000 were obtained which is in accordance with the fish BCFs described in the literature (Table 17), even though a broad range of BCF values was described for both substances. For pentachlorobenzene, the in vitro and in vivo fish tests both provided BCFs >5000. In this case, the HYBIT test, carried out by Landrum et al. (2004) according to a different testing protocol compared to HYBIT, resulted in a BCF of 1913. However, no lipid values are available to allow lipid normalisation of the BCF value to 3% TL w/w. It can be only speculated whether lipid normalisation would have led to a higher HYBIT BCF value. The importance of lipid normalisation in the calculation of HYBIT BCFs was demonstrated in the interlaboratory test (OECD 2024b).

Table 17: Lipid partitioning (Scheme 1): log K_{OW} 4.5 – 6.

	Tier 1		Tier 2	Tier 3		Tier 4
Chemical	Consolidated log K_{OW}	Consolidated log K_{OA}	Mean fish QSAR BCF	IVIVE BCF	HYBIT BCF	Experimental fish BCF range
Chlorpyrifos	4.89	9.47	934	509	1163	40 - 3162
Pyrene	4.91	8.75	1618	144	3050	50 - 1351
Triclosan	4.91	11.08	1553	146	N/A	1 - 1995
Octamethylcyclo-tetrasiloxane (D4)	4.98	4.13	1679	1935§	test failed	5012 - 14900
Trifloxystrobin	5.08	11.2	579	672	568	372 - 537
Pentachloro-benzene	5.11	6.53	6603	5499**	1913§§	3981 - 22909
Methoxychlor	5.25	10.6	2623	1928 (S9) 3404 – 3765 (HEP)	9396	621 - 8300
o-terphenyl	5.58	8.80	3555	test failed#	6086	501 - 12882

* No lipid normalisation

** No biotransformation determinable

§ R^2 in 1 out of 3 runs > 0.85

§§ No lipid normalisation, median value

No first order depletion rate determined

Even though substances of moderate to high hydrophobicity are within the applicability domain of the NAMs, in two cases the performance of tests failed. The HYBIT test with the difficult-to-test compound D4 was not technically feasible. In the case of o-terphenyl, no first order depletion rate could be determined in the fish IVIVE approach.

Generally, the results show that in several cases expert judgement on the results of the NAMs is required to avoid misinterpretation with regard to the chemicals' bioaccumulation potential and that it is necessary to consider all available data in a WoE approach.

6.2.3 log K_{OW} > 6 (superhydrophobic compounds)

For three substances—benzo[a]pyrene, PCB 153, and UV-234—consolidated log K_{OW} values above 6 were determined (Table 18). For PCB 153, the QSAR, HYBIT, IVIVE and in vivo fish BCF values are far above 5000, no biotransformation was determined in the in vitro study, thus the results are consistent with each other.

For benzo[a]pyrene, the mean fish QSAR BCF is estimated at 3669, which corresponds well to the HYBIT BCF of 3911. However, both, the IVIVE and the experimental fish BCF are below 1000. ECHA has concluded in its SVHC support document (ECHA 2016b), which also lists other bioaccumulation data, that "The bioaccumulation potential of B[a]P strongly varies with the organism´s ability to metabolise PAHs. In general, fish and other vertebrates are able to efficiently metabolise B[a]P due to the presence of Cytochrome P450-like enzymes, leading to a low to moderate BCF. In contrast, many invertebrates (in particular mussels and crustaceans) are lacking those enzymes suitable for metabolising B[a]P and, therefore, very high BCFs were observed in the range from 3680 to 140700. Thus, it is concluded that B[a]P is a bioaccumulative substance" (ECHA 2016b). In the case of this chemical, the in vitro and the in vivo fish test give consistent results but are not conservative. Here, it was necessary to take into consideration all available data beyond that which is considered in this project to be able to conclude on the bioaccumulative properties of the substance, and to understand the reason why the substance does not accumulate as highly in fish as it does in invertebrates. For UV-234, in vitro BCF values exceeding 5000 were obtained. No biotransformation was determined. Both in vivo studies with *Hyalella azteca* and fish show results below 2000. For *Hyalella azteca*, the BCF was determined in this project, see description in A.2.3. Similar to previous studies (Schlechtriem, 2022), the experiment met with some difficulties, given the high log K_{OW} of this substance. The fish BCF study was conducted by the registrant, the summary of the report (ECHA registration dossier) is according to the guideline. Due to systematic errors in the test performance, as explained in detail in the next paragraph, the fish and *Hyalella azteca* in vivo studies, especially the exposure phases, are considered not reliable for regulatory assessment. Recalculation resulted in BCF values exceeding 5000.

Given the high hydrophobicity of these three chemicals, the applicability and reliability of the different test systems must be critically evaluated in light of the obtained results. Each test should be assessed individually using expert judgement to determine whether the outcomes are plausible and valid. Although all studies were conducted in compliance with the respective test guidelines, particular attention in both HYBIT and in vivo fish tests should be paid to the uptake and depuration rate constants, the extended time required to reach steady state, and the influence of TOC on the freely dissolved fraction of the test substance. Likewise, the applicability of the in vitro test for assessing superhydrophobic substances should be critically evaluated. For substances with a log K_{OW} > 6, a specific assessment may be required, if testing is not validly feasible. Alternative strategies for the bioaccumulation assessment of such substances are currently being developed within another UBA-funded project [Project No. (FKZ) 3723 64 405 0].

Table 18: Lipid partitioning (Scheme 1): log K_{OW} > 6 (superhydrophobic compounds).

	Tier 1		Tier 2	Tier 3		Tier 4
Chemical	Consolidated log K_{OW}	Consolidated log K_{OA}	Mean fish QSAR BCF	IVIVE BCF	HYBIT BCF	Experimental fish BCF range
Benzo[a]-pyrene	6.07	11.0	3669	787	3911	3 - 631
PCB 153	7.08	10.0	195265	13596*	155373	8913 - 740000
UV-234	7.68	15.32	5761	7464*	1670	1286

* No biotransformation determinable

6.3 Assessment of bioaccumulation in air-breathing organisms (Scheme 2)

For substances that are evaluated as non-bioaccumulative in aquatic organisms, it should be further investigated whether there could be a potential for bioaccumulation in air-breathing organisms. If log K_{OW} > 2, log K_{OA} > 5, and the estimated biohalf-life is greater than 4 days in rats, an additional assessment is suggested for terrestrial bioaccumulation potential (Scheme 2, and ECHA 2023a). The determination of a biohalf-life will have to be discussed elsewhere.

The criterion for no aquatic bioaccumulation and "log K_{OW} > 2" and "log K_{OA} > 5" were fulfilled for a couple of substances, as shown in Table 15. Among them are, e.g., azoxystrobin, fluorene, diclofenac, β-HCH, phenanthrene, chlorpyrifos, terbutryn, and others. As this small example underlines, the additional screening criterium biohalf-life, or another suitable one, will be necessary to reduce the data load for additional testing to perform a bioaccumulation assessment in air-breathing organisms. Table 19 summarises all IVIVE BMF values that could be derived as part of this project. In vitro depletion data used for the extrapolations were either taken from the literature or were generated as part of this project (*cf.* chapter A.9).

IVIVE BMF values > 1 were extrapolated for β-HCH, methoxychlor and UV-234. The substance β-HCH did not show accumulative behaviour in aquatic organisms, but in air-breathing organisms (Kelly 2007a). As Table 19 demonstrates, the IVIVE BMF data supports the monitoring findings. As a conclusion, the application of NAM methods alone could help in identifying substances that do not bioaccumulate in organisms of the aquatic environment, but in air-breathing organisms. There are no additional substances in the exemplary sample set that show bioaccumulation tendencies in air-breathing organisms, but not in water breathing organisms. Methoxychlor is already rated B to vB based on aquatic data, a IVIVE BMF of 1.5 confirms the assessment. UV-234 and D4 were tested in the rat in vitro depletion assay to provide further data. The UV-234 data in the aquatic assessment scheme range from BCF = 1286 (Fish in vivo BCF) to BCF = 7464 (IVIVE BCF). In three out of four runs of the rat in vitro depletion assay no significant depletion of the test compound could be detected. Only in one run biotransformation could be determined. Both IVIVE based BMFs are above 1 and could thus indicate a potential for biomagnification in air-breathing organisms. It should, however, be further investigated, if a biotransformation of this compound is possible for hepatic rat material to further refine the data. The rat in vitro depletion assay with D4 could not be conducted, due to technical problems.

The IVIVE BMFs listed in Table 19 were calculated using the B-compass Rat model (Krause & Goss 2021) for hepatocytes as described in chapter A.5.6, since the IVIVE model published by Lee et al. 2022 can be used for extrapolations based on S9 data, only. The B-compass Rat (Krause & Goss 2021) model allows an extrapolation of both, S9 substrate depletion data and hepatocyte

substrate depletion data. Compared to the Lee et al. (2022) model, the B-compass Rat model extrapolations result in considerably higher IVIVE BMF estimates. Compared to a pyrene IVIVE BMF value of 0.022 calculated by the Lee et al. 2022 model, the B-compass Rat model calculates a IVIVE BMF of 2.6 based on the same depletion rate (*cf.* chapter A.9).

Table 19: Lipid partitioning (Scheme 1a): Bioaccumulation in air-breathing mammals.

	Tier 1		Tier 2	Tier 3		Tier 4	Air-breathers
Chemical	Consolidated log K_{OW}	Consolidated log K_{OA}	Mean fish QSAR BCF	IVIVE BCF	HYBIT BCF	Experimental fish BCF range	IVIVE BMF (rat)
β-HCH	3.87	7.59	814	367	887	293 - 1460	24
Diclofenac	4.35	12.9	376	0 – 944*§	3-73	5-9	0 – 0.12*
Pyrene	4.91	8.75	1618	144	3050	50 - 1351	0.773
D4	4.98	4.13	1679	1935	test failed	5012 - 14900	test failed
Chlorpyrifos	4.89	9.47	934	509	1163	40 - 3162	0.08
Methoxychlor	5.25	10.6	2623	1928 (S9) 3404 – 3765 (HEP)	9396	621 - 8300	1.5
UV-234	7.68	15.32	5761	7464**	1670	1286	4 - 33**

Grey-shaded cells: IVIVE model not applicable to ionisable substances. Values given nonetheless to provide best- and worst-case scenarios

* Log D values for ionised and neutral species of the substance inserted into IVIVE model; IVIVE BCF = 0 obtained with log D value for ionised species

** No biotransformation determinable

§ Low data quality

6.4 Substances not predominantly accumulating in lipid

6.4.1 Ionisable substances

Three ionisable substances (fluoxetine, diclofenac, pentachlorophenol) are part of the group of case study chemicals (Table 20). Log D values of 0.93, 0.7 and 2.45 are reported for fluoxetine, diclofenac and pentachlorophenol, respectively. The corresponding consolidated log K_{OW} values are 4.33, 4.35 and 4.81 (*cf.* chapter 5.4). The calculated QSAR BCF values were below 1000 (*cf.* chapter 5.3) which was confirmed by HYBIT BCF and experimental fish BCF values. Only IVIVE BCF values could reach values above 1000 and 2000. In this case, Log D values of the neutral species of the substance were inserted into the IVIVE model. If log D values of the ionised chemical species were inserted, the IVIVE BCFs were below values of 100. However, the B-compass Fish IVIVE model is not applicable to ionisable substances. Values are presented nonetheless to provide best- and worst-case scenarios (*cf.* chapter A.4.4). The results for the three ionisable substances do not correlate well with either the HYBIT or the fish BCF which underlines that an IVIVE model designed for ionisable chemicals should be used, ideally. These models require validation and the availability of reliable input data. For pentachlorophenol, the results of the HYBIT test (BCF 132) need to be considered with caution, since no lipid normalisation had been applied.

Table 20: Testing ionisable substances.

		Tier 1		Tier 2	Tier 3		Tier 4
Chemical	Conso-lidated log K_{OW}	Log D @ pH	Conso-lidated log K_{OA}	Mean fish QSAR BCF	IVIVE BCF	HYBIT BCF	Experimental fish BCF range
Fluoxetine	4.33	0.93*	10	323	0 – 1040	19 - 348	13 - 330
Diclofenac	4.35	0.7*	12.9	376	0 – 944[§]	3-73	5-9
Pentachloro-phenol	4.81	4.16 @ 5.5** 2.45 @ 7.4**	9.31	835	14 – 2976[§§]	132[#]	195 - 479

Grey-shaded cells: IVIVE model not applicable to ionisable substances. Values given nonetheless to provide best- and worst-case scenarios. Log D values for ionised and consolidated log K_{OW} values for neutral species of the substance inserted into IVIVE model; IVIVE model not within AD

*Reported as "log D ionic" in Köhler et al. 2023

**Calculated by the ACD/Labs Percepta Platform - PhysChem Module, retrieved from ChemSpider.com (https://www.chemspider.com/Chemical-Structure.967.html, last accessed 14.10.2025)

[§] Slope quality rates as "low" by authors

[§§] No biotransformation determinable

[#] not lipid normalised

6.4.2 PFAS

PFAS are regarded as proteinophilic compounds. Two PFAS, PFOS and GenX, are part of the group of case study chemicals (Table 21). For GenX a consolidated log K_{OW} of 3.26 was determined. The calculated QSAR BCF was clearly below 100 indicating a low bioaccumulation potential. This was confirmed by further studies in tier 3, for both HYBIT and fish in vitro BCFs, and in vivo fish studies in tier 4. The testing of PFOS resulted in QSAR BCF and HYBIT BCF values indicating no bioaccumulation potential. However, a fish BCF test (tier 4) provided clear indications for a high bioaccumulation potential of the chemical (BCF > 2000). In contrast to the other tests, no in vitro data are available for PFOS. This chemical is known to accumulate highly in air-breathers which led to its inclusion in the Stockholm Convention. Further investigations with a broader set of sorptive chemicals are required. With regard to the bioaccumulation of neutral organic chemicals, HYBIT BCFs tend to be higher compared to the respective fish BCF values. For chemicals such as PFAS this might be different and may require a critical revision of the assessment scheme with respect to the use of the NAMs. Specific physicochemical properties of PFAS necessitate the use of a specific mechanistic bioaccumulation model to evaluate the bioaccumulation potential of PFAS (*cf.* chapter 4.1.2).

Table 21: Sorption: Testing PFAS.

	Tier 1		Tier 2	Tier 3		Tier 4
Chemical	Consolidated log K_{OW}	Consolidated log K_{OA}	Mean fish QSAR BCF	IVIVE BCF	HYBIT BCF	Experimental fish BCF range
GenX	3.26	5.09	5.4	91	7**	8
PFOS	4.42	5.81	91.3	N/A	302*	255 - 5400

Grey-shaded cells: IVIVE model not applicable to ionisable substances/surfactants.

* Median BCF_k value, not lipid normalised

** Not lipid normalised

6.4.3 Surfactants

The bioaccumulation potential of surfactants or compounds that exhibit amphiphilic properties was also assessed in this project by the example of lauric acid. Due to the specific phase behaviour of surfactants and their full ionisation under test conditions, these substances are outside the applicability domain of most tools used to calculate log K_{OW}. For the non-ionised form of lauric acid, a high consolidated log K_{OW} of 4.84 was calculated. However, the QSAR-BCF of 238 indicated a lack of bioaccumulation. This was confirmed by HYBIT and the results of an in vivo fish test (Table 22). Also, in this case, further investigations with a broader set of chemicals that exhibit amphiphilic properties are recommended to allow a critical evaluation of the proposed assessment scheme.

Table 22: Substances with amphiphilic characteristics: Testing the surfactant lauric acid.

	Tier 1		Tier 2	Tier 3		Tier 4
Chemical	Consolidated log K_{OW}	Consolidated log K_{OA}	Mean fish QSAR BCF	IVIVE BCF	HYBIT BCF	Experimental fish BCF range
Lauric acid	4.84	8.51	238	N/A	6	255

6.5 Conclusions

The tiered assessment according to Scheme 1 demonstrates that NAMs, including QSARs, in vitro assays (OECD TG 319A/B), and the HYBIT test (OECD TG 321), can provide a scientifically robust framework for evaluating the bioaccumulation potential of organic chemicals across a wide range of hydrophobicities. The predictive value of log K_{OW} alone was confirmed, reliably identifying non-bioaccumulative (non-B) chemicals without the need for further testing for substances with low hydrophobicity ($K_{OW} < 4$), however, as shown for anthracene and phenanthrene, chemicals with a log K_{OW} of < 4.5 can still lead to results in tier 3-4 indicating a bioaccumulation potential.

For moderately to highly hydrophobic compounds, ECHA has suggested a threshold BCF of 1200 and 3000 for the HYBIT for confirmatory testing. A similar approach with an adjusted threshold BCF could be used for the in vitro test. The exact value of this threshold BCF remains to be determined. Considering the inconsistencies between HYBIT, in vitro, and in vivo fish BCFs

which were evident for several compounds (e.g. pentachlorophenol, pyrene), the use of such threshold BCF values seems sensible. BCF values derived by these NAMs that exceed the B and vB values can be directly used for decision-making.

The variability in the test results, as shown in Table 15, underlines the necessity of expert judgement and, in some cases, confirmatory testing for borderline cases. These further highlight the need for complementary NAMs approaches or higher-tier evaluations as well as the consideration of all available data, if any additional data is available.

Superhydrophobic compounds are difficult to test in flow-through in vivo fish or HYBIT tests. The results of this project highlight both the relevance and the limitations of current NAMs for testing such compounds. While certain outcomes (e.g. PCB 153) were consistent across methods, divergent results for others (e.g. benzo[a]pyrene, UV-234) point to methodological challenges for NAMs such as extended times to steady state or the influence of TOC, and the limited applicability of in vitro tests for such compound tests. Dedicated assessment strategies for these substances are still under development.

The case studies with ionisable, sorptive or amphiphilic substances (e.g. PFAS, lauric acid) show that the scope of application of current NAMs is not universal, which also applies to the use of these substances in in vivo fish tests. Even though the currently available data is still limited, the current data results suggests that in vitro assays with hepatocytes or S9 fractions can provide valuable data on the metabolic biotransformation of chemicals in air-breathing organisms. These methods support IVIVE-based estimations of bioaccumulation potential and could reduce animal testing. Further work is needed to validate hepatocyte and liver S9 fraction assays, extend their use to, e.g., rat, human, and bird, validate the required IVIVE models and to incorporate them into regulatory guidance. Development of standardised test guidelines would be welcome.

Overall, the study confirms that NAMs are well suited for reliably assessing bioaccumulation concern under defined conditions, thereby reducing the need for in vivo/vertebrate/animal testing. At the same time, the findings underscore the importance of critical evaluation, expert judgement, and, where necessary, complementary testing for borderline, superhydrophobic, or technically challenging substances. NAMs therefore represent a promising but not yet universally applicable replacement for the fish bioaccumulation test.

7 Concluding reflections and future pathways

The transformation of chemical regulation over recent decades has been driven by a growing understanding of environmental complexity, ethical imperatives, and technological innovation. This is also clearly reflected in the assessment of bioaccumulation – a key criterion for determining the environmental persistence and toxicity of chemical substances. The integrated assessment strategy developed in this project reflects this broader evolution, providing a structured and adaptable framework that addresses the shortcomings of traditional methods while embracing new approaches fit for 21st-century regulatory science.

7.1 Responding to historical limitations

Historically, the assessment of bioaccumulation has centred almost exclusively on fish bioconcentration studies, most notably those conducted under OECD Test Guideline 305. While this test has been instrumental in identifying hazardous substances and guiding chemical regulation globally, it has also become clear that its scope is inherently limited. Not all substances behave predictably in aquatic organisms and test conditions are not always good enough to produce satisfactory results (e.g. for superhydrophobic compounds). Moreover, reliance on vertebrate animal testing has come under increased scrutiny due to both ethical considerations and the growing availability of scientifically valid non-animal methods.

The emergence of complex substance classes such as PFAS, benzotriazoles, and amphiphilic chemicals has further revealed the limitations of models which have been commonly applied. These substances often fall outside the applicability domains of conventional testing, resulting in inconclusive or misleading assessments.

In response to these challenges, the integrated assessment scheme outlined in this study offers a tiered, evidence-based, and animal-sparing approach that integrates physicochemical profiling, computational models, and advanced laboratory testing, with the goal of producing robust, interpretable, and harmonised bioaccumulation assessments across regulations.

7.2 Structure and logic of the integrated assessment scheme

The integrated assessment scheme is built on a five-tier framework, each level introducing increasing levels of complexity and resource investment, while prioritising methods that avoid or minimise animal use:

Tier 1 – Physicochemical Screening

The first tier focuses on fundamental physicochemical parameters, such as log K_{OW} and log K_{OA}, but also including other parameters such as log D, as relevant for the respective substances. These values serve as screening tools to quickly eliminate substances with clearly low bioaccumulation potential. Importantly, the project demonstrated that both measured and modelled values for log K_{OW} and log K_{OA} can vary widely—often exceeding one log unit. No single method proved consistently reliable, necessitating a consensus-based approach to ensure robust assessments. For hydrophilic chemicals, the bioaccumulation potential in aquatic organisms is considered low, and no further testing is required under any European legislation. This recommendation is backed by the project data. Superhydrophobic substances (log $K_{OW} > 6$), however, require specialised evaluation, which is currently being developed in follow-up projects (FKZ 3723 64 4050 “Alternative strategies for the assessment of bioaccumulation of superhydrophobic substances”).

Tier 2 – QSAR Modelling

In Tier 2 QSAR models are used to estimate bioconcentration factors (BCFs). While QSARs provide a fast, cost-effective alternative to animal testing, the study highlighted significant variability between different models, even within their respective applicability domains. As with log K_{OW}, a consensus modelling approach was recommended. For all substances with predicted BCFs below 500, the BCF values derived by the HYBIT, in vitro and in vivo fish tests were all also below 500. Nevertheless, these QSAR methods require further validation for regulatory use.

Tier 3 – New Approach Methodologies (NAMs)

Tier 3 introduces more advanced test systems, particularly the in vitro hepatic clearance assays (OECD TG 319A/B) and the *Hyalella azteca* bioconcentration test (OECD TG 321). The decision which of these tests should be used, should be taken based on substance properties and feasibility of conducting a valid test, which ultimately has an influence on resources needed. These methods represent a significant step forward, offering viable alternatives to fish testing for many substances. In practice, the tests showed good correlation with fish BCF data for low- to mid-hydrophobic substances. However, discrepancies can arise in mid- to highly hydrophobic or complex chemicals. In this case a follow-up with the other test is recommended. Where both approaches yield conflicting results, expert judgement or higher-tier testing is needed. Characteristics of the HYBIT and in vitro tests are compared in Table 23.

Table 23: Characteristics of OECD TGs 319A/B and 321 to display informative values and options of these two NAMs.

	OECD TG 319A/B	OECD 321
Endpoint	Biotransformation rate (in vitro) In silico extrapolation to IVIVE BCF possible	In vivo BCF (Invertebrate)
Organism	Rainbow trout Mid to upper trophic position	*Hyalella azteca* Lower trophic position
Target	Parent substance	Parent substance or TRR
Provided insight	Specific insight into hepatic biotransformation of the test substance. Can be used to compare biotransformation of isomers	Uptake and elimination of target substance (and all metabolites with ^{14}C-testing). No systematic insight into different compartments of the organism.
Applicable to superhydrophobic substances	Substance dependent Current evaluation in UBA project 3723 64 405 0	Substance dependent Current evaluation in UBA project 3723 64 405 0
Applicable to ionisable substances	Depletion assay: Yes IVIVE: Work in progress, soon to be published	Yes Testing at different pH values possible
Applicable to substances that do not (predominantly)	Depletion assay: Yes IVIVE: Work in progress	Exposure: Yes Data modelling and normalisation: Work in progress

	OECD TG 319A/B	OECD 321
accumulate in neutral lipids		
Metabolite pattern analysis	Possible (requires additional effort)	Possible (requires additional effort)
Metabolism pathways	Can be deduced (requires additional effort)	Can be deduced (requires additional effort)
Accumulation of biotransformation products	Can be extrapolated from metabolite structure (requires additional effort)	Can be co-determined using radiolabel testing (requires additional effort)

Tier 4 – In Vivo Confirmation

The final tier includes in vivo fish testing (OECD TG 305) but should only be employed when lower-tier tests provide inconclusive outcomes. Importantly, the project emphasises that such testing should be considered a last resort, and only under clear regulatory necessity.

Tier 5 – Monitoring data and additional information

If additional information, including environmental monitoring data and non-guideline studies, is available, this should always be included in the evaluation in a WoE approach. This can be especially helpful in cases where in vivo data remain ambiguous. This has been shown for various substances in this project, where additional data was needed to draw conclusions on the true bioaccumulation potential of substances that did not show high accumulation in laboratory tests.

Bioaccumulation in air-breathing vertebrates

Bioaccumulation in air-breathing vertebrates has long been neglected in regulatory frameworks. This project incorporated rat hepatocyte in vitro assays to simulate hepatic metabolism in air-breathing organisms. While promising, more research is needed to develop an accepted test guideline for rat hepatocytes and liver S9 fractions, possibly also for human and bird hepatocytes and S9 fractions and integrate them into regulatory guidance.

7.3 Testing for special substance classes

The project paid special attention to substances that present unique assessment challenges:

- **Superhydrophobic compounds**, such as UV-234, displayed anomalous BCF behaviour, probably due to limited bioavailability. The project confirmed earlier findings that these compounds may be misclassified as non-bioaccumulative under traditional testing. For these substances, customised assessment strategies are in development.
- **For ionisable compounds**, the log D should be used instead of log K_{OW}. Laboratory testing should be conducted at the pH where the substance is ionised to the lowest degree. For the in vitro test, specific models have to be used that account for the ionised fraction.
- **PFAS and other sorptive substances** pose difficulties due to their tendency to adhere to surfaces rather than partition into lipid phases. Current NAMs showed limited applicability, with PFOS being the only PFAS tested. Further work is needed to adapt test methods for these chemicals.

- **Amphiphilic substances**, such as surfactants, challenge log K_{OW}-based methods due to ionisation and phase behaviour. The integrated assessment scheme includes guidance for calculating log K_{OW} under variable pH conditions and recommends dual assessments for neutral and ionised forms, as well as suggesting the use of alternative partitioning factors.

7.4 Scientific and regulatory impact

The integrated assessment scheme provides the basis to evolve the existing assessment frameworks in a way that maintains the same level of protection as current EU regulatory approaches, while fostering innovation and reducing reliance on animal testing. Its adoption would confirm a paradigm shift in how regulatory authorities assess chemical behaviour in the environment. It reflects the growing scientific consensus that a one-size-fits-all approach to bioaccumulation is neither effective nor ethically sustainable. By integrating computational modelling, in vitro systems, and non-vertebrate organisms, the integrated assessment scheme supports a move toward mechanistic, evidence-based assessment frameworks that are adaptable across regulatory domains. This transformation also supports the implementation of the 3Rs in chemical regulation. The integrated assessment scheme minimises animal use by filtering out low-risk substances early and using NAMs to confirm results before resorting to vertebrate tests only for those cases where other approaches remain inconclusive. The approach offers not only ethical advantages but also cost savings, faster testing timelines, and increased flexibility for regulators and industry stakeholders alike.

From a regulatory perspective, the integrated assessment scheme presented here lays a foundation for an assessment procedure that can be integrated into the different EU substance regulations. The project's stakeholder workshop in Berlin highlighted widespread interest in this goal and identified priorities for method standardisation, data sharing, and training.

7.5 Remaining challenges and recommendations

Despite its strengths, the integrated assessment scheme also reveals areas where further work is essential:

- **Data Transparency**: Improved databases and curation standards are needed to ensure that key input parameters (e.g., log K_{OW}, log K_{OA}, pKa) are reliable.
- **Model Validation**: Both QSAR and in vitro models require broader validation across more diverse chemical classes to improve regulatory acceptability.
- **Expert Judgement**: For difficult-to-test substances, expert-driven WoE approaches must be formally incorporated into regulatory guidance.
- **Method Development**: New assays tailored to amphiphilic and non-lipid binding compounds are urgently needed. Also, development of an in vitro test guideline for rat, bird and/or human S9 fractions and hepatocytes, or for other organisms which are specific indicators in ecosystems, would be highly welcome. This should also include validation of the related IVIVE models to derive bioaccumulation endpoints. Specific models for specific substance classes such as ionisable compounds should be developed.
- **Secondary poisoning:** The HYBIT BCF should be integrated into the assessment of secondary poisoning. The IVIVE BCF derived from an OECD 319A/B test can be used in the same way as the fish BCF is used.

- **Terrestrial Integration**: Expansion of test strategies for air-breathing organisms will be key to assessing broader ecological impacts.

7.6 Summary

The integrated assessment scheme presented in this project provides a comprehensive and forward-looking framework for bioaccumulation assessment that can be integrated into the different EU chemical regulations by the relevant committees. It enables regulators to move beyond outdated reliance on vertebrate testing, supports the assessment of the broad variety of chemical substances, and promotes cross-regulatory coherence.

Built on a flexible five-tier structure, the integrated assessment scheme emphasises early screening, use of in silico data and NAMs, to ensure minimal animal testing. The study shows that NAMs can effectively exclude bioaccumulation concerns under certain conditions, reducing reliance on animal tests. Yet, for challenging substances, critical review and complementary methods remain essential.

In the context of growing environmental pressures, technological capability, and societal expectations, the integrated assessment scheme offers a concept for the next generation of chemical safety assessment—one that is evidence-based, efficient, and ethically sound. Its adoption will not only enhance the scientific quality of regulatory decisions but also contribute to a more sustainable and protective approach to environmental health.

8 Acknowledgements

The project was supported by an international group of experts whose valuable contribution to the success of this project is gratefully acknowledged:

Aude Kienzler, EFSA

Doris Hirmann, ECHA

Eric Verbruggen, RIVM

Eszter Simon, BAFU

Rhys Whomsley, EMA

Heike Laue, Givaudan

Maike Habekost, BASF

Sandrine Charles, Université de Lyon

Our sincere appreciation is extended to the UBA internal advisory group whose guidance and support were instrumental to this project:

Thomas Gräff

Ines Prutz

Gabriele Treu

Peter von der Ohe

We sincerely thank the many colleagues who actively participated in the stakeholder workshop in Berlin and contributed valuable feedback and suggestions that helped improve this report. Finally, we would like to thank the UBA project coordinators, Caren Rauert and Daniel Frein, for their unwavering dedication and exceptional support throughout the project.

9 References

ACD/Labs (2023). Physicochemical Property Calculator. https://www.acdlabs.com.

Ahlers J, Nendza M, Schwartz D (2019). Environmental hazard and risk assessment of thiochemicals. Application of integrated testing and intelligent assessment strategies (ITS) to fulfil the REACH requirements for aquatic toxicity. Chemosphere 214:480-490. https://doi.org/10.1016/j.chemosphere.2018.09.082.

Allendorf F, Berger U, Goss K-U, Ulrich N (2019). Partition coefficients of four perfluoroalkyl acid alternatives between bovine serum albumin (BSA) and water in comparison to ten classical perfluoroalkyl acids. Environmental Science: Processes & Impacts 21(11):1852-1863. https://doi.org:10.1039/C9EM00290A.

ARC Arnot Research and Consulting Inc. (2023). EAS-E Suite (Ver.0.97 - BETA, release June, 2023). https://www.eas-e-suite.com.

ARChem (2010) SPARC Performs Automated Reasoning in Chemistry. https://www.archemcalc.com/sparc.html.

Armitage JM, Gobas FA (2007). A terrestrial food-chain bioaccumulation model for POPs. Environ Sci Technol. 41(11):4019-4025. https://doi.org/10.1021/es0700597.

Armitage JM, Arnot JA, Wania F, Mackay D (2013). Development and evaluation of a mechanistic bioconcentration model for ionogenic organic chemicals in fish. Environ Toxicol Chem 32(1):115–128. https://doi.org/10.1002/etc.2020.

Armitage JM, Erickson RJ, Luckenbach T, Ng CA, Prosser RS, Arnot JA, Schirmer K, Nichols JW (2017). Assessing the bioaccumulation potential of ionizable organic compounds: Current knowledge and research priorities. Environ toxicol and chem, 36(4): 882–897. https://doi.org/10.1002/etc.3680.

Armitage J, Toose L, Embry M, Foster KL, Hughes L, Sangion A, Arnot JA (2021). The Bioaccumulation Assessment Tool (BAT) Version 2.02. ARC Arnot Research and Consulting Inc., Toronto, ON, Canada. https://www.arnotresearch.com.

Arnot JA, Gobas FAPC (2003). A generic QSAR for assessing the bioaccumulation potential of organic chemicals in aquatic food webs. QSAR CombiSci 22:337-345.

Arnot JA, Toose L, Armitage JM, Embry M, Sangion A, Hughes L (2023). A weight of evidence approach for bioaccumulation assessment. Integr Environ Assesst and Manag19:1235–1253. https://doi.org/: 10.1002/ieam.4583.

Balls M, Combes RD (2005). The early history of the Three Rs concept. Alternatives to Laboratory Animals, 33(1): 3–16.

Baskaran S, Lei YD, Wania F (2021). A Database of Experimentally Derived and Estimated Octanol–Air Partition Ratios (KOA). Journal of Physical and Chemical Reference Data 50(4):043101. https://doi.org/10.1063/5.0059652.

Baskaran S, Wania F (2023). Applications of the octanol-air partitioning ratio: A critical review. Environmental Science: Atmospheres. https://doi.org/10.1039/D3EA00046J.

Benfenati E (2010). The CAESAR project for in silico models for the REACH legislation. Chemistry Central Journal 4 (1):I1. https://doi.org/10.1186/1752-153X-4-S1-I1.

Bintein S, Devillers J, Karcher W (1993). Nonlinear dependence of fish bioconcentration on n-octanol/water partition coefficient. SAR QSAR EnvironRes 1:29-39.

BIOVIA (2023). COSMOtherm. https://www.3ds.com/products-services/biovia/products/molecular-modeling-simulation/solvation-chemistry/biovia-cosmotherm/.

Bischof I, Köster J, Segner H, Schlechtriem C (2016). Hepatocytes as in vitro test system to investigate metabolite patterns of pesticides in farmed rainbow trout and common carp: Comparison between in vivo and in vitro and across species. Comp Biochem Physiol C Toxicol Pharmacol187:62-73. https://doi.org/10.1016/j.cbpc.2016.05.003.

Bischof I, Arnot JA, Jürling H, Knipschild G., Schlechtriem C, Schauerte A, Segner H (2022). In vitro biotransformation assays using fish liver cells: Comparing rainbow trout and carp hepatocytes; Frontiers in Toxicology; Volume 4. https://doi.org/0.3389/ftox.2022.1021880.

Black SR, Nichols JW, Fay KA, Matten SR, Lynn SG (2021). Evaluation and comparison of in vitro intrinsic clearance rates measured using cryopreserved hepatocytes from humans, rats, and rainbow trout. Toxicology 457:152819. https://doi.org/10.1016/j.tox.2021.152819.

Böhm L, Schlechtriem C, Düring, RA (2016). Sorption of Highly Hydrophobic Organic Chemicals to Organic Matter Relevant for Fish Bioconcentration Studies. Environ Sci Technol 50: 8316–8323. https://doi.org/10.1021/acs.est.6b01778.

Brandmaier S, Tetko IV, Öberg T (2012). An evaluation of experimental design in QSAR modelling utilizing the k-medoid clustering. Journal of Chemometrics 26(10):509-517. https://doi.org/10.1002/cem.2459.

Brooke DN, Crookes MJ, Gray D, Robertson S (2009). Environmental Risk Assessment Report: Octamethylcyclotetrasiloxane, Environment Agency (UK), ISBN: 978-1-84911-031-0.

Burkhard LP (2000). Estimating dissolved organic carbon partition coefficients for nonionic organic chemicals. Environ Sci Technol 34:4663–4668. https://doi.org/10.1021/es001269l.

Butt CM, Muir DC, Mabury SA (2010). Biotransformation of the 8:2 fluorotelomer acrylate in rainbow trout. 2. In vitro incubations with liver and stomach S9 fractions. Environ toxicol chem 29(12): 2736–2741. https://doi.org/10.1002/etc.348.

Cantu MA, Gobas FAPC (2023). Bioaccumulation of Linear Siloxanes in Fish. Environ Toxicol Chem 43 (1):42-51. https://doi.org/10.1002/etc.5760

Carson, R (1962). Silent Spring. Boston: Houghton Mifflin, 1962.

Castren M,Oikari, A (1983). Optimal assay conditions for liver UDP-glucuronosyltransferase from the rainbow trout, *Salmo gairdneri*. Comp Biochem Physiol Part C 76:365–369.

Chemaxon (2023). JChem. https://chemaxon.com.

Chemaxon (2024). Chemicalize. https://chemaxon.com.

Chen Y, Hermens JL, Jonker MT, Arnot JA, Armitage JM, Brown T, Nichols JW, Fay KA, Droge ST (2016). Which Molecular Features Affect the Intrinsic Hepatic Clearance Rate of Ionizable Organic Chemicals in Fish? Environ Sci Technol 50(23):12722–12731. https://doi.org/10.1021/acs.est.6b03504.

Connell DE, Schüürmann G (1988). Evaluation of Various Molecular Parameters as Predictors of Bioconcentration in Fish. Ecotoxicol Environ Saf 15:324-335.

Connell DW, Hawker DW (1988). Use of polynomial expressions to describe the bioconcentration of hydrophobic chemicals by fish. Ecotoxicol Environ Saf 16:242-257.

Connors KA, Du B, Fitzsimmons PN, Chambliss CK, Nichols JW, Brooks BW (2013). Enantiomer-specific in vitro biotransformation of select pharmaceuticals in rainbow trout (*Oncorhynchus mykiss*). Chirality 25(11):763–767. https://doi.org/10.1002/chir.22211.

DataWarrior (2013). cLogP from Property Explorer. https://www.openmolecules.org.

de Wolf W, Comber M, Douben P, Gimeno S, Holt M, Léonard M, Lillicrap A, Sijm D, van Egmond R, Weisbrod A, Whale G (2007). Animal use replacement, reduction, and refinement: development of an integrated testing strategy for bioconcentration of chemicals in fish. Integr Environ Assess Manag 3(1):3–17. https://doi.org/10.1897/1551-3793(2007)3[3:AURRAR]2.0.CO;2.

Dimitrov SD, Dimitrova NC, Walker JD, Veith GD, Mekenyan OG (2003). Bioconcentration potential predictions based on molecular attributes - an early warning approach for chemicals found in humans, birds, fish and wildlife. QSAR CombiSci 22:58-68.

Dong X, Xu Z, Zhao H, Wu D, Qu B, Liu S, Xiao B (2025). Predictive modeling and interpretability analysis of bioconcentration factors for organic chemicals in fish using machine learning. Environmental Pollution 377:126323. https://doi.org/10.1016/j.envpol.2025.126323.

Droge STJ, Scherpenisse P, Arnot JA, Armitage JM, McLachlan MS, Ohe PCvd, Hodges G (2021). Screening the baseline fish bioconcentration factor of various types of surfactants using phospholipid binding data. Environmental Science: Processes & Impacts 23(12):1930-1948. https://doi:10.1039/D1EM00327E.

Ebert A, Ackermann J, Goss KU, (2023a). Mechanistic modeling of the bioconcentration of (super)hydrophobic compounds in *Hyalella azteca*. Environ Sci Pollut Res 50257–50268. https://doi.org/10.1007/s11356-023-25827-7.

Ebert RU, Kühne R, Schüürmann G (2023b). Octanol/Air Partition Coefficient – A General-Purpose Fragment Model to Predict Log K_{oa} from Molecular Structure. Environ Sci Technol 57 (2):976-984. https://doi.org/10.1021/acs.est.2c06170.

EC (2001). Directive 2001/83/EC of the European Parliament and of the Council of 6 November 2001 on the Community code relating to medicinal products for human use.

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), establishing a European Chemicals Agency, amending Directive 1999/45/EC and repealing Council Regulation (EEC) No 793/93 and Commission Regulation (EC) No 1488/94 as well as Council Directive 76/769/EEC and Commission Directives 91/155/EEC, 93/67/EEC, 93/105/EC and 2000/21/EC.

EC (2008). Regulation (EC) no 1272/2008 of the European Parliament and of the Council of 16 December 2008 on classification, labeling and packaging of substances and mixtures, amending and repealing Directives 67/548/EEC and 1999/45/EC, and amending Regulation (EC) No 1907/2006.

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.

EC (2012). DG Sanco Working Document on „Evidence needed to identify POP, PBT and vPvB Properties for Pesticides“ 2012 rev 3.

ECHA (2008). Support Document for identification of anthracene as a substance of very high concern, Member State Committee, Adopted on 8 October 2008.

ECHA (2016a). Guidance on IR&CSA – Chapter R.16: Environmental exposure assessment – Version 3.0 (Feb 2016). Reference: ECHA-16-G-03-EN.

ECHA (2016b). Support Document for identification of Benzo[def]chrysene (Benzo[a]pyrene) as a substance of very high concern because of its carcinogenic (Article 57 A), mutagenic (Article 57 B), toxic for reproduction (Article 57 C), persistent, bioaccumulative, and toxic (Article 57 D) and very persistent and very bioaccumulative (Article 57 E) properties, Member State Committee, Adopted on 27 May 2016.

ECHA (2017a). Guidance on BPR – Volume IV Environment - Assessment and Evaluation (Parts B+C), Version 2.0, Oct 2017, Reference/ISBN: ECHA-17-G-23-EN. https://echa.europa.eu/documents/10162/2324906/bpr_guidance_ra_vol_iv_part_b-c_en.pdf/e2622aea-0b93-493f-85a3-f9cb42be16ae?t=1513159347576.

ECHA (2017b). Guidance on information requirements and chemical safety assessment Chapter R.7a: Endpoint specific guidance. https://www.op.europa.eu/en/publication-detail/-/publication/182ca463-f427-11e7-9d43-01aa75ed71a1/language-en.

ECHA (2017c). Read-Across Assessment Framework (RAAF). https://www.echa.europa.eu/documents/10162/13628/raaf_en.pdf.

ECHA (2018). Support Document for identification of phenanthrene as a substance of very high concern because of its vPvB (Article 57 E) properties, Member State Committee, Adopted on 12 December 2018.

ECHA (2022a). Bioaccumulation assessment of air-breathing mammals: a discussion paper. https://www.echa.europa.eu/documents/10162/17228/bioaccumulation_assessment_of_air_breathing_mammals_en.pdf/56de6276-06e9-9eed-a7dd-a75336fda71b?t=1669388928484.

ECHA (2022b). Guidance on the Biocidal Products Regulation: Volume IV: Environment - Part A: Information Requirements. ISBN: 978-92-9468-112-6. https://doi: 10.2823/339326.

ECHA (2023a). Guidance on Information Requirements and Chemical Safety Assessment Chapter R.11: PBT/vPvB assessment Version 4.0. https://www.echa.europa.eu/documents/10162/13632/information_requirements_r11_en.pdf/a8cce23f-a65a-46d2-ac68-92fee1f9e54f.

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

ECHA (2023c). Registration Dossier 1,1'-(ethane-1,2-diyl)bis[pentabromobenzene]. https://www.echa.europa.eu/registration-dossier/-/registered-dossier/15001/1/1.

ECHA (2023d). Report on the European Chemicals Agency's "New Approach Methodologies Workshop: Towards an Animal Free Regulatory System for Industrial Chemicals" 31 May – 1 June 2023, Helsinki, Finland. ECHA-23-R-09-EN. ISBN: 978-92-9468-299-4. https://doi: 10.2823/7494.

ECHA (2023e). Registered substances (2023) https://www.echa.europa.eu/information-on-chemicals/registered-substances.

ECHA (2024). Guidance on the Application of the CLP Criteria, Parts 4 and 5 Version 4.0 – Nov 2024. ECHA-24-G-05-EN Cat. Number: ED-01-24-015-EN-N. ISBN: 978-92-9468-411-0. https://doi: 10.2823/4895485.

eChemPortal (2023). https://www.echemportal.org/echemportal/.

EFSA (2009). Guidance Document on Risk Assessment for Birds and Mammals. EFSA Journal 7(12):1438. https://doi.org/10.2903/j.efsa.2009.1438.

EFSA (2011). Conclusion on the peer review of the pesticide risk assessment of the active substance prochloraz. EFSA Journal 9(7):2323. https://doi:10.2903/j.efsa.2011.2323.

EFSA (2013). PPR Panel (EFSA Panel on Plant Protection Products and their Residues), 2013. Guidance on tiered risk assessment for plant protection products for aquatic organisms in edge-of-field surface waters. EFSA Journal 2013;11(7):3290. https://doi:10.2903/j.efsa.2013.3290.

EMA (2015). Guideline on the assessment of persistent, bioaccumulative and toxic (PBT) or very persistent and very bioaccumulative (vPvB) substances in veterinary medicinal products. EMA/CVMP/ERA/52740/2012.

EMA (2016). Guideline on environmental impact assessment for veterinary medicinal products in support of the VICH guidelines GL6 and GL38, EMA/CVMP/ERA/418282/2005-Rev.1.

EMA (2020). Regulatory science to 2025: Strategic reflection. EMA/110706/2020. https://www.ema.europa.eu/en/documents/regulatory-procedural-guideline/regulatory-science-2025-strategic-reflection_en.pdf.

EMA (2024). Guideline on ERA of medicinal products for human use – Rev.1 – Corr – effective 01/09/2024. EMEA/CHMP/SWP/4447/00 Rev.1 – Corr.

Endo S, Goss KU (2014). Applications of Polyparameter Linear Free Energy Relationships in Environmental Chemistry. Environ Sci Technol 48(21):12477-12491. https://doi:10.1021/es503369t.

European Parliament and Council (2010). Directive 2010/63/EU on the protection of animals used for scientific purposes. Official Journal of the European Union, L276, 33–79.

EU (European Union) (2019). Regulation (EU) 2019/6 of the European Parliament and of the Council of 11 December 2018 on veterinary medicinal products and repealing Directive 2001/82/EC. Official Journal of the European Union, L 4, 7 January 2019, 43–167.

Fay KA, Fitzsimmons PN, Hoffman AD, Nichols JW (2014a). Optimizing the use of rainbow trout hepatocytes for bioaccumulation assessments with fish. Xenobiotica 44(4): 345–351. https://doi.org/10.3109/00498254.2013.845704.

Fay KA, Mingoia RT, Goeritz I, Nabb DL, Hoffman AD, Ferrell BD, Peterson HM, Nichols JW, Segner H, Han X (2014b). Intra- and interlaboratory reliability of a cryopreserved trout hepatocyte assay for the prediction of chemical bioaccumulation potential. Environ Sci Technol 48(14):8170–8178. https://doi.org/10.1021/es500952a.

Fay KA, Fitzsimmons PN, Hoffman AD, Nichols JW (2017). Comparison of trout hepatocytes and liver S9 fractions as in vitro models for predicting hepatic clearance in fish. Environ Toxicol Chem 36:463-471. https://doi.org/10.1002/etc.3572.

Fliszkiewicz B, Sajdak M (2023). Fragments quantum descriptors in classification of bio-accumulative compounds. Journal of Molecular Graphics and Modelling 125:108584. https://doi.org/10.1016/j.jmgm.2023.108584.

Fu W, Franco A, Trapp S (2009). Methods for estimating the bioconcentration factor (BCF) of ionizable organic chemicals. Environ Toxicol Chem 28:1372-1379.

Fu Q, Rösch A, Fedrizzi D, Vignet C, Hollender J (2018). Bioaccumulation, Biotransformation, and Synergistic Effects of Binary Fungicide Mixtures in *Hyalella azteca* and *Gammarus pulex*: How Different/Similar are the Two Species? Environ Sci Technol 52(22): 13491–13500. https://doi.org/10.1021/acs.est.8b04057.

Fu Q, Fedrizzi D, Kosfeld V, Schlechtriem C, Ganz V, Derrer S, Rentsch D, Hollender J (2020). Biotransformation Changes Bioaccumulation and Toxicity of Diclofenac in Aquatic Organisms. Environ Sci Technol 54(7): 4400–4408. https://doi.org/10.1021/acs.est.9b07127.

Glüge J, Escher BI, Scheringer M (2022). How error-prone bioaccumulation experiments affect the risk assessment of hydrophobic chemicals and what could be improved. Integr Environ Assess Manag 19(3):792-803. https://doi.org/10.1002/ieam.4714.

Gobas FA, de Wolf W, Burkhard LP, Verbruggen E, Plotzke K (2009), Revisiting bioaccumulation criteria for POPs and PBT assessments. Integr Environ Assess Manag 5(4):624-37. https://doi: 10.1897/IEAM_2008-089.1.

Gobas FAPC, Moore MM, Hermens JLM, Arnot JA (2006). Bioaccumulation reality check. SETAC Globe 7(5):40-41.

Gobas FAPC, Kelly B, Arnot J (2003). Quantitative Structure Activity Relationships for Predicting the Bioaccumulation of POPs in Terrestrial Food-Webs. QSAR CombiSci 22:329-336. https://doi.org/10.1002/qsar.200390022.

Goss KU, Brown TN, Endo S (2013). Elimination half-life as a metric for the bioaccumulation potential of chemicals in aquatic and terrestrial food chains. Environ Toxicol Chem 32(7):1663-1671. https://doi.org/10.1002/etc.2229.

Gramatica P, Papa E (2005). An Update of the BCF QSAR Model Based on Theoretical Molecular Descriptors. QSAR CombiSci 24(8):953-960. https://doi.org/10.1002/qsar.200530123.

Gust KA, Steevens JA, Noyes PD, Zhang Y, Szostek B, Newsted JL, Gaworecki KM (2025). Bioconcentration, maternal transfer, and toxicokinetics of perfluorooctane sulfonate (PFOS) in a multigenerational zebrafish (*Danio rerio*) exposure. Environ Toxicol Chem. Advance online publication. https://doi.org/10.1093/etojnl/vgae033.

Habig WH, Pabst MJ, Jakoby WB (1974). Glutathione S-transferases. The first enzymatic step in mercapturic acid formation. Journal of Biological Chemistry 249:7130–7139.

Halder M, Lostia A, Kienzler A (2018). EURL ECVAM Fish In Vitro Intrinsic Clearance Database. European Commission, Joint Research Centre (JRC) [Dataset] PID: http://data.europa.eu/89h/jrc-eurl-ecvam-fish-in-vitro-intr-clear-db.

Han X, Nabb DL, Mingoia RT, Yang CH (2007). Determination of xenobiotic intrinsic clearance in freshly isolated hepatocytes from rainbow trout (Oncorhynchus mykiss) and rat and its application in bioaccumulation assessment. Environ Sci Technol 41(9):3269–3276. https://doi.org/10.1021/es0626279.

Han X, Nabb DL, Yang CH, Snajdr SI, Mingoia RT (2009). Liver microsomes and S9 from rainbow trout (*Oncorhynchus mykiss*): comparison of basal-level enzyme activities with rat and determination of xenobiotic intrinsic clearance in support of bioaccumulation assessment. Environ Toxicol Chem 28(3):481–488. https://doi.org/10.1897/08-269.1.

Hansch C, Fujita T (1964). ρ-σ-π-Analysis. A method for the correlation of biological activity and chemical structure. J Am Chem 86:1616-1626.

Hansch C, Leo AJ (1979). Substituent constants for correlation analysis in chemistry and biology. Wiley, New York, NY.

Hodges G, Eadsforth C, Bossuyt B, Bouvy A, Enrici M-H, Geurts M, Kotthoff M, Michie E, Miller D, Müller J, Oetter G, Roberts J, Schowanek D, Sun P, Venzmer J (2019). A comparison of log Kow (n-octanol–water partition coefficient) values for non-ionic, anionic, cationic and amphoteric surfactants determined using predictions and experimental methods. Environmental Sciences Europe 31(1):1. https//doi.org/10.1186/s12302-018-0176-7.

Hodgeson E, Levi PE (1994). Introduction to Biochemical Toxicology. Appleton & Lange, Norwalk, CT, USA.

Istituto di Ricerche Farmacologiche Mario Negri IRCCS (2023). VEGA. Laboratory of Environmental Chemistry and Toxicology. https://www.vegahub.eu/.

Jonker M, van der Haijden S (2007a). BCF hydrophobicity cut-off: Truth or fake? SETAC Europe 17th Annual Meeting, Porto:RA04B-01.

Jonker M, van der Haijden S (2007b). Bioconcentration factor hydrophobicity cut-off: An artificial phenomenon reconstructed. Environ Sci Technol 41(21):7363-7369.

Kamlet MJ, Doherty RM, Abraham MH, Marcus Y, Taft RS (1988). Linear Solvation Energy Relationships. An Improved Equation for Correlation and Prediction of Octanol/Water Partition Coefficients of Organic Nonelectrolytes (Including Strong Hydrogen Bond Donor Solutes). J Phys Chem 92:5244-5255.

Kelly BC, Ikonomou MG, Blair JD, Morin AE, Gobas FAPC (2007a). Food web-specific biomagnification of persistent organic pollutants. Science 317(5835):236-239. https://doiorg/10.1126/science.1138275. Erratum in: Science 318(5847):44.

Kelly BC, Ikonomou MG, Gobas FAPC (2007b). Bioaccumulation of persistent organic pollutants in lichen–caribou–wolf food chains of Canada's central and western Arctic. Environ Sci Technol 41(2): 548–555. https://doi.org/10.1021/es062236+.

Kelly BC, Sun JM, McDougall MRR, Sunderland EM, Gobas FAPC (2024). Environ Sci Technol 58(40):17828-17837. https://doi.org/10.1021/acs.est.4c02134.

Klamt A, Schüürmann G (1993). COSMO: a new approach to dielectric screening in solvents with explicit expressions for the screening energy and its gradient. Journal of the Chemical Society, Perkin Transactions 2(5):799-805. https://doi.org/10.1039/P29930000799.

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. Ecological Informatics 63:101285. https://doi.org/10.1016/j.ecoinf.2021.101285.

Köhler HR, Gräff T, Schweizer M, Blumhardt J, Burkhardt J, Ehmann L, Hebel J, Heid C, Kundy L, Kuttler J, Malusova M, Moroff FM, Schlösinger AF, Schulze-Berge P, Panagopoulou EI, Damalas DE, Thomaidis NS, Triebskorn R, Maletzki D, Kühnen U, von der Ohe PC (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 Research 235:119864. ISSN 0043-1354, https://doi.org/10.1016/j.watres.2023.119864.

Kolanczyk RC, Saley MR, Serrano JA, Daley SM, Tapper MA (2023). PFAS Biotransformation Pathways: A Species Comparison Study. Toxics 11(1):74. https://doi.org/10.3390/toxics11010074.

Könemann H, van Leeuwen K (1980). Toxicokinetics in fish: Accumulation and elimination of six chlorobenzene by guppies. Chemosphere 9:3-19.

Kosfeld V, Fu Q, Ebersbach I, Esser D, Schauerte A, Bischof I, Hollender J, Schlechtriem C (2020). Comparison of Alternative Methods for Bioaccumulation Assessment: Scope and Limitations of In Vitro Depletion Assays with Rainbow Trout and Bioconcentration Tests in the Freshwater Amphipod *Hyalella azteca*. Environ Toxicol Chem 39(9):1813-1825. https://doi.org/10.1002/etc.4791.

Kowalska D, Sosnowska A, Zdybel S, Stepnik M, Puzyn T (2024). Predicting bioconcentration factors (BCFs) for per- and polyfluoroalkyl substances (PFAS). Chemosphere 364:143146. https://doi.org/10.1016/j.chemosphere.2024.143146.

Krause S, Goss KU (2020). Comparison of a simple and a complex model for BCF prediction using in vitro biotransformation data. Chemosphere 256:127048.

Krause S, Goss KU (2021). B-compass Rat_Kow v1.0. Leipzig, Germany, Helmholtz Centre for Environmental Research.

Kroll A, von der Ohe PC, Köhler HR, Sellier O, Junghans M (2024). Aquatic thresholds for ionisable substances, such as diclofenac, should consider pH-specific differences in uptake and toxicity. Science of The Total Environment 908:168222. https://doi.org/10.1016/j.scitotenv.2023.168222.

Kropf C, Begnaud F, Gimeno S, Berthaud F, Debonneville C, Segner H (2020). In vitro biotransformation assays using liver S9 fractions and hepatocytes from rainbow trout (*Oncorhynchus mykiss*): Overcoming challenges with difficult to test fragrance chemicals. Environ Toxicol Chem 39:2396–2408. https://doi.org/10.1002/etc.4872.

Kuehr S, Kaegi R, Maletzki D, Schlechtriem C (2021). Testing the bioaccumulation potential of manufactured nanomaterials in the freshwater amphipod *Hyalella azteca*. Chemosphere, 263: 127961. https://doi.org/10.1016/j.chemosphere.2020.127961.

Kühne R, Ebert RU, Schüürmann G (2009). Chemical domain of QSAR models from atom-centered fragments. J Chem Inf Model 49:2660-2669.

Ladd MA, Fitzsimmons PN, Nichols JW (2016). Optimization of a UDP-glucuronosyltransferase assay for trout liver S9 fractions:activity enhancement by alamethicin, a pore-forming peptide. Xenobiotica 46: 1066.

Landrum PF, Scavia D (1983). Influence of Sediment on Anthracene Uptake, Depuration, and Biotransformation by the Amphipod *Hyalella azteca*. Canadian Journal of Fisheries and Aquatic Sciences 40(3):298-305. https://doi.org/10.1139/f83-044.

Landrum PF, Steevens JA, Gossiaux DC, McElroy M, Robinson S, Begnoche L, Chernyak S, Hickey J (2004). Time-dependent lethal body residues for the toxicity of pentachlorobenzene to *Hyalella azteca*. Environ Toxicol Chem 23(5):1335–1343. https://doi.org/10.1897/03-164.

Larisch W, Goss KU (2017). Modelling Oral Uptake of Hydrophobic and Super-Hydrophobic Chemicals in Fish, Environ. Sci.: Processes Impacts 20(1): 98–104.

Larisch W, Goss KU (2018). Uptake, distribution and elimination of chemicals in fish – Which physiological parameters are the most relevant for toxicokinetics? Chemosphere 210: 1108-1114.

Laue H, Gfeller H, Jenner KJ, Nichols JW, Kern S, Natsch A (2014). Predicting the bioconcentration of fragrance ingredients by rainbow trout using measured rates of in vitro intrinsic clearance. Environ Sci Technol 48(16): 9486–9495. https://doi.org/10.1021/es500904h.

Laue H, Hostettler L, Badertscher RP, Jenner KJ, Sanders G, Arnot JA, Natsch A (2020). Examining Uncertainty in In Vitro-In Vivo Extrapolation Applied in Fish Bioconcentration Models. Environ Sci Technol 54(15):9483–9494. https://doi.org/10.1021/acs.est.0c01492.

Laue H, Hostettler L, Jenner KJ, Sanders G, Natsch A (2023). Bioconcentration Assessment in Fish Based on In Vitro Intrinsic Clearance: Predictivity of an Empirical Model Compared to In Vitro–In Vivo Extrapolation Models; Environ Sci Technol 57(36): 13325-13335. https://doi.org/10.1021/acs.est.3c02216.

Laville N, Aït-Aïssa S, Gomez E, Casellas C, Porcher JM (2004). Effects of human pharmaceuticals on cytotoxicity, EROD activity and ROS production in fish hepatocytes. Toxicology 196(1–2):41–55. https://doi.org/10.1016/j.tox.2003.11.002.

Lee JH, Landrum PF, Koh CH (2002). Toxicokinetics and time-dependent PAH toxicity in the amphipod *Hyalella azteca*. Environ Sci Technol 36(14):3124-30. https://doi.org/10.1021/es011201l.

Lee YS, Lee DH, Delafoulhouze M, Otton SV, Moore MM, Kennedy CJ, Gobas FAPC (2014). In vitro biotransformation rates in fish liver S9: effect of dosing techniques. Environ Toxicol Chem 33(8):1885–1893. https://doi.org/10.1002/etc.2636.

Lee YS, Lo JC, Otton SV, Moore MM, Kennedy CJ, Gobas FAPC (2017). In vitro to in vivo extrapolation of biotransformation rates for assessing bioaccumulation of hydrophobic organic chemicals in mammals. Environ Toxicol Chem 36(7):1934–1946. https://doi.org/10.1002/etc.3718.

Lee YS, Cole TR, Jhutty MS, Cantu MA, Chee B, Stelmaschuk SC, Gobas FAPC (2022). Bioaccumulation Screening of Neutral Hydrophobic Organic Chemicals in Air-Breathing Organisms Using In Vitro Rat Liver S9 Biotransformation Assays. Environ Toxicol Chem 41(10):2565-2579. https://doi.org/10.1002/etc.5439.

Li L, Zhang Z, Men Y, Baskaran S, Sangion A, Wang S, Arnot JA, Wania F (2022). Retrieval, Selection, and Evaluation of Chemical Property Data for Assessments of Chemical Emissions, Fate, Hazard, Exposure, and Risks. ACS Environmental Au 2(5):376-395. https://doi.org/10.1021/acsenvironau.2c00010.

Lo JC, Allard GN, Otton SV, Campbell DA, Gobas FA (2015). Concentration dependence of biotransformation in fish liver S9: Optimizing substrate concentrations to estimate hepatic clearance for bioaccumulation assessment. Environ Toxicol Chem 34(12):2782–2790. https://doi.org/10.1002/etc.3117.

Lombardo A, Roncaglioni A, Benfenati E, Nendza M, Segner H, Fernández A, Kühne R, Franco A, Pauné E, Schüürmann G (2014). Integrated Testing Strategy (ITS) for bioaccumulation assessment under REACH. Environ Int 69:40–50.

Louisse J, Alewijn M, Peijnenburg AACM, Cnubben NHP, Heringa MB, Coecke S, Punt A (2020). Towards harmonization of test methods for in vitro hepatic clearance studies. Toxicol In Vitro:104722. https://doi.org/10.1016/j.tiv.2019.104722.

Lu X, Tao S, Cao J, Dawson RW (1999). Prediction of fish bioconcentration factors of nonpolar organic pollutants based on molecular connectivity indices. Chemosphere 39(6):987-999. https://doi.org/10.1016/S0045-6535(99)00020-X.

Luo W, Zhou G, Zhu Z, Yuan Y, Ke G, Wei Z, Gao Z, Zheng H (2024). Bridging Machine Learning and Thermodynamics for Accurate pKa Prediction. JACS Au 4(9):3451-3465. https://doi.org/10.1021/jacsau.4c00271

Lyman WJ, Reehl WF, Rosenblatt DH (1982). Handbook of Chemical Property Estimation Methods: Environmental Behavior of Organic Compounds. McGraw-Hill. New York, USA.

Mackay D (1982). Correlation of bioconcentration factors. Environ Sci Technol 16(5):274-278.

Mackay D, Celsie AKD, Parnis JM (2015). The evolution and future of environmental partition coefficients. Environmental Reviews 24(1):101-113. https://doi.org/10.1139/er-2015-0059.

Mandel J (1984). The statistical analysis of experimental data. Dover Publications Inc.. NY, USA.

Mansouri K, Cariello NF, Korotcov A, Tkachenko V, Grulke CM, Sprankle CS, Allen D, Casey WM, Kleinstreuer NC, Williams AJ (2019). Open-source QSAR models for pKa prediction using multiple machine learning approaches. Journal of Cheminformatics 11(1):60. https://doi.org/10.1186/s13321-019-0384-1.

Mansouri K, Grulke CM, Judson RS, Williams AJ (2018). OPERA models for predicting physicochemical properties and environmental fate endpoints. Journal of Cheminformatics 10(1):10. https://doi.org/10.1186/s13321-018-0263-1.

Martin JW, Mabury SA, Solomon KR, Muir DC (2003). Bioconcentration and tissue distribution of perfluorinated acids in rainbow trout (*Oncorhynchus mykiss*). Environ Toxicol Chem 22(1):196–204.

Martin YC (2009). Let's not forget tautomers. Journal of Computer-Aided Molecular Design 23(10):693-704. https://doi.org/10.1007/s10822-009-9303-2.

McLachlan MS, Czub G, MacLeod M, Arnot JA (2011). Bioaccumulation of organic contaminants in humans: a multimedia perspective and the importance of biotransformation. Environ Sci Technol 45(1):197-202. https://doi.org/10.1021/es101000w.

McLachlan MS, Ebert A, Armitage JM, Arnot JA, Droge ST (2023). A Framework for Understanding the Bioconcentration of Surfactants in Fish. Environmental Science: Processes and Impacts 25 (7):1238–1251. https://doi.org/10.1039/D3EM00070B.

Medina EIS, Linke S, Sundmacher K (2021). Prediction of Bioconcentration Factors (BCF) using Graph Neural Networks. In: Türkay M, Gani R (eds). Computer Aided Chemical Engineering, vol 50. Elsevier, The Netherlands. pp 991-997. https://doi.org/10.1016/B978-0-323-88506-5.50153-4.

Meylan WM, Howard PH, Boethling RS, Aronson D (1999). Improved method for estimating bioconcentration/bioaccumulation factor from octanol/water partition coefficient. Environ Toxicol Chem 18:664-672.

Mingoia RT, Glover KP, Nabb DL, Yang CH, Snajdr SI, Han X (2010). Cryopreserved hepatocytes from rainbow trout (*Oncorhynchus mykiss*): a validation study to support their application in bioaccumulation assessment. Environ Sci Techn 44(8): 3052–3058. https://doi.org/10.1021/es903909g.

Molinspiration Cheminformatics free web services (2023). miLogP from Molinspiration Property Calculator. https://www.molinspiration.com.

Mueller C, Trapp S, Polesel F, Kuehr S, Schlechtriem C (2020). Biomagnification of ionisable organic compounds in rainbow trout *Oncorhynchus mykiss*. Environmental Sciences Europe 32:159. https://doi.org/10.1186/s12302-020-00443-1.

Müller M, Nendza M (2007). Literature study: Effects of molecular size and lipid solubility on bioaccumulation potential. https://www.umweltbundesamt.de/sites/default/files/medien/2674/dokumente/bcf.pdf.

Nabb DL, Mingoia RT, Yang CH, Han Y (2006). Comparison of basal level metabolic enzyme activities of freshly isolated hepatocytes from rainbow trout (*Oncorhynchus mykiss*) and rat, Aquatic Toxicology 80:52–59.

Nakamura Y, Yamamoto H, Sekizawa J, Kondo T, Hirai N, Tatarazako N (2008). The effects of pH on fluoxetine in Japanese medaka (*Oryzias latipes*): acute toxicity in fish larvae and bioaccumulation in juvenile fish. Chemosphere 70(5):865–873. https://doi.org/10.1016/j.chemosphere.2007.06.089.

National Research Council (2007). Toxicity Testing in the 21st Century: A Vision and a Strategy. Washington, DC: National Academies Press.

Navo CD, Jiménez-Osés G (2021). Computer Prediction of pKa Values in Small Molecules and Proteins. ACS Medicinal Chemistry Letters 12(11):1624-1628. https://doi.org/10.1021/acsmedchemlett.1c00435.

Neely WB, Branson DR, Blau GE (1974). Partition coefficient to measure bioconcentration potential of organic chemicals in fish. Environ Sci Technol 8:1113-1115.

Nendza M (1991). QSARs of bioconcentration: validity assessment of log Pow/log BCF correlations. Bioaccumulation in Aquatic Systems: In: Nagel R, Loskill R. Bioaccumulation in Aquatic Systems: Contributions to the Assessment. VCH Verlagsgesellschaft, Weinheim, Germany, pp. 43-66.

Nendza M, Hermens JLM (1995). Properties of chemicals and estimation methodologies. In: van Leeuwen CJ, Hermens JLM (eds). Risk Assessment of Chemicals: An Introduction. Kluwer Academic Publishers, Dordrecht, The Netherlands, pp 239-292.

Nendza M (1998). Structure-activity relationships in environmental sciences. Chapman & Hall, London, Great Britain.

Nendza M, Aldenberg T, Benfenati E, Benigni R, Cronin MTD, Escher S, Fernandez A, Gabbert S, Giralt F, Hewitt M, Hrovat M, Jeram S, Kroese D, Madden JC, Mangelsdorf I, Rallo R, Roncaglioni A, Rorije E, Segner H, Simon-Hettich B, Vermeire TG (2010). Data quality assessment for in silico methods: A survey of approaches and needs. In: Cronin MTD, Madden JC (eds). In Silico Toxicology: Principles and applications. Royal Society of Chemistry, Cambridge, UK, pp 59-117.

Nendza M, Gabbert S, Kühne R, Lombardo A, Roncaglioni A, Benfenati E, Benigni R, Bossa C, Strempel S, Scheringer M, Fernández A, Rallo R, Giralt F, Dimitrov S, Mekenyan O, Bringezu F, Schüürmann G (2013). A comparative survey of chemistry-driven in silico methods to identify hazardous substances under REACH. Regulatory Toxicology and Pharmacology 66(3):301-314.

Nendza M, Kosfeld V, Schlechtriem C (2025). Consolidated octanol/water partition coefficients: Combining multiple estimates from different methods to reduce uncertainties in log K_{OW}. Environmental Sciences Europe 37, Article 44. https://doi.org/10.1186/s12302-025-01072-2.

Nentwig G (2007). Effects of pharmaceuticals on aquatic invertebrates. Part II: the antidepressant drug fluoxetine. Archives of environmental contamination and toxicology 52(2):163–170. https://doi.org/10.1007/s00244-005-7190-7.

Nichols JW, Huggett DB, Arnot JA, Fitzsimmons PN, Cowan-Ellsberry CE (2013a). Toward improved models for predicting bioconcentration of well-metabolized compounds by rainbow trout using measured rates of in vitro intrinsic clearance. Environ Toxicol Chem 32:1611-1622. https://doi.org/10.1002/etc.2219.

Nichols JW, Hoffman AD, ter Laak TL, Fitzsimmons PN (2013b). Hepatic Clearance of 6 Polycyclic Aromatic Hydrocarbons by Isolated Perfused Trout Livers: Prediction From In Vitro Clearance by Liver S9 Fractions, Toxicological Sciences, Volume 136(2): 359–372. https://doi.org/10.1093/toxsci/kft219.

Nichols J, Fay K, Bernhard MJ, Bischof I, Davis J, Halder M, Hu J, Johanning K, Laue H, Nabb D, Schlechtriem C, Segner H, Swintek J, Weeks J, Embry M (2018a). Reliability of In Vitro Methods Used to Measure Intrinsic Clearance of Hydrophobic Organic Chemicals by Rainbow Trout: Results of an International Ring Trial. Toxicological Sciences 164 (2):563-575. https://doi.org/10.1093/toxsci/kfy113.

Nichols JW, Ladd MA, Fitzsimmons PN (2018b). Measurement of kinetic parameters for biotransformation of polycyclic aromatic hydrocarbons by trout liver S9 fractions: Implications for bioaccumulation assessment. Appl In Vitro Toxicol 4(4):365-378. https://doi.org/10.1089/aivt.2017.0005.

Oakes KD, Coors A, Escher BI, Fenner K, Garric J, Gust M, Knacker T, Küster A, Kussatz C, Metcalfe CD, Monteiro S, Moon TW, Mennigen JA, Parrott J, Péry AR, Ramil M, Roennefahrt I, Tarazona JV, Sánchez-Argüello P, Ternes TA, … Servos MR (2010). Environmental risk assessment for the serotonin re-uptake inhibitor fluoxetine: Case study using the European risk assessment framework. Integrated Environmental Assessment and Management, 6 Suppl, 524–539. https://doi.org/10.1002/ieam.77.

OECD (1995). Test No. 107: Partition Coefficient (n-octanol/water): Shake Flask Method. https://doi.org/10.1787/9789264069626-en.

OECD (2007). Guidance document on the validation of (quantitative) structure-activity relationships [(Q)SAR] models, vol OECD Environment Health and Safety Publications. Series on Testing and Assessment No. 69. OECD, Paris, France.

OECD (2010). Test No. 317: Bioaccumulation in Terrestrial Oligochaetes. OECD Guidelines for the Testing of Chemicals, Section 3. OECD Publishing, Paris, France. https://doi.org/10.1787/9789264090934-en.

OECD (2012). Test No. 305: Bioaccumulation in Fish: Aqueous and Dietary Exposure. OECD Publishing, Paris, France. https://doi.org/10.1787/9789264185296-en.

OECD (2018a). Test No. 319A: Determination of in vitro intrinsic clearance using cryopreserved rainbow trout hepatocytes (RT-HEP). OECD Guidelines for the Testing of Chemicals, Section 3, OECD Publishing, Paris, France. https://doi.org/10.1787/9789264303218-en.

OECD (2018b). Test No. 319B: Determination of in vitro intrinsic clearance using rainbow trout liver S9 sub-cellular fraction (RT-S9). OECD Guidelines for the Testing of Chemicals, Section 3. OECD Publishing, Paris, France. https://doi.org/10.1787/9789264303232-en.

OECD (2018c). Guidance document on the determination of in vitro intrinsic clearance using cryopreserved hepatocytes (RT-HEP) or liver S9 sub-cellular fractions (RT-S9) from rainbow trout and extrapolation to in vivo intrinsic clearance. Series on Testing and Assessment. No. 280. ENV/JM/MONO (2018)12. OECD Publishing, Paris, France.

OECD (2022a). Test No. 123: Partition Coefficient (1-octanol/water): Slow-Stirring Method. OECD Publishing, Paris, France. https://doi.org/10.1787/9789264015845-en.

OECD (2022b). Test No. 117: Partition Coefficient (n-octanol/water), HPLC Method. OECD Publishing, Paris, France. https://doi.org/10.1787/9789264069824-en.

OECD (2023a). (Q)SAR Assessment Framework: Guidance for the regulatory assessment of (Quantitative) Structure - Activity Relationship models, predictions, and results based on multiple predictions. OECD Publishing, Paris, France. https://www.oecd.org/chemicalsafety/risk-assessment/qsar-assessment-framework.pdf.

OECD (2023b). QSAR Toolbox https://www.oecd.org/chemicalsafety/risk-assessment/oecd-qsar-toolbox.htm.

OECD (2024a). Test No. 321: *Hyalella azteca* Bioconcentration Test (HYBIT), OECD Guidelines for the Testing of Chemicals, Section 3, OECD Publishing, Paris, France. https://doi.org/10.1787/8ac30c4e-en.

OECD (2024b). Multi-Laboratory Ring Trial to Support Development of OECD Test Guideline 321 on *Hyalella azteca* Bioconcentration Test (HYBIT), OECD Series on Testing and Assessment, No. 398, OECD Publishing, Paris, France. https://doi.org/10.1787/cfb97663-en.

OECD (2024c). Case study on the Use of Integrated Approaches for Testing and Assessment (IATA) for Bioaccumulation - Ninth Review Cycle (2023), OECD Series on Testing and Assessment, No. 404, OECD Publishing, Paris, France. https://doi.org/10.1787/8bc8ad6f-en.

Opperhuizen A, Wijbenga A (1984). Bioaccumulation of super-lipophilic chemicals in fish. Toxicol Environ Chem 7:173–189. https://doi.org/10.1080/02772248409357024.

Padberg K (2024). The Influence of pH on the Bioaccumulation of Ionic Substances in the Freshwater Amphipod *Hyalella azteca*. Master thesis. University of Münster, Germany.

Papa E, Dearden JC, Gramatica P (2007). Linear QSAR regression models for the prediction of bioconcentration factors by physicochemical properties and structural theoretical molecular descriptors. Chemosphere 67(2):351-358. https://doi.org/10.1016/j.chemosphere.2006.09.079.

PerkinElmer Informatics (2023). ClogP from ChemDraw (Version 19.1.1.21). https://www.biobyte.com/.

Perrin DD, Dempsey B, Serjeant EP (1981). pKa Prediction for Organic Acids and Bases. Springer, Dordrecht, The Netherlands. https://doi.org/10.1007/978-94-009-5883-8.

Pfeifer S, Pflegel P, Borchert HH (1984). Grundlagen der Biopharmazie. Verlag Chemie, Weinheim, Germany.

Pliška V, Testa B, Waterbeemd Hvd (eds) (1996). Lipophilicity in Drug Action and Toxicology. VCH Verlagsgesellschaft mbH, Weinheim, Germany. https://doi.org/10.1002/9783527614998.

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. Journal of Hazardous Materials 479:135725. https://doi.org/10.1016/j.jhazmat.2024.135725.

Potter TD, Barrett EL, Miller MA (2021). Automated Coarse-Grained Mapping Algorithm for the Martini Force Field and Benchmarks for Membrane–Water Partitioning. Journal of Chemical Theory and Computation 17(9):5777-5791. https://doi.org/10.1021/acs.jctc.1c00322.

Qin W, Henneberger L, Huchthausen J, König M, Escher BI (2023). Role of bioavailability and protein binding of four anionic perfluoroalkyl substances in cell-based bioassays for quantitative in vitro to in vivo extrapolations. Environment International 173:107857. https://doi.org/10.1016/j.envint.2023.107857.

Raths J, Kühr S, Schlechtriem C (2020). Bioconcentration, metabolism and spatial distribution of ^{14}C-labelled laurate in the freshwater amphipod *Hyalella azteca*. Environ Toxicol Chem 39:310–322. https://doi.org/10.1002/etc.4623.

Rekker RF (1977). The hydrophobic fragmental constant. Elsevier Scientific, New York, NY, USA.

Richardson SJ, Bai A, Kulkarni AA, Moghaddam MF (2016). Efficiency in Drug Discovery: Liver S9 Fraction Assay As a Screen for Metabolic Stability. Drug metabolism letters 10(2):83–90. https://doi.org/10.2174/1872312810666160223121836.

Roberts SC, Noyes PD, Gallagher EP, Stapleton HM (2011). Species-specific differences and structure-activity relationships in the debromination of PBDE congeners in three fish species. Environ Sci Technol 45(5):1999–2005. https://doi.org/10.1021/es103934x.

Royal Society of Chemistry (2018). ChemSpider. https://www.chemspider.com.

Russell WMS, RL Burch RL (1959). The Principles of Humane Experimental Technique. Methuen, USA.

Schlechtriem C, Böhm L, Bebon R, Bruckert HJ, Düring RA (2017). Fish bioconcentration studies with column-generated analyte concentrations of highly hydrophobic organic chemicals. Environ Toxicol Chem 36:906–916. https://doi.org/10.1002/etc.3635.

Schlechtriem C, Kampe S, Bruckert H-J, Bischof I, Ebersbach I, Kosfeld V, Kotthoff M, Schäfers C, L'Haridon J (2019). Bioconcentration studies with the freshwater amphipod *Hyalella azteca*: are the results predictive of bioconcentration in fish? Environmental Science and Pollution Research 26(2):1628-1641. https://doi.org/10.1007/s11356-018-3677-4.

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. Environmental Sciences Europe 24:13. https://doi.org/10.1186/2190-4715-24-13.

Schlechtriem C, Kühr S, Müller C (2022) Development of a bioaccumulation test using *Hyalella azteca*. Texte | 134/2022, Umweltbundesamt, Dessau-Roßlau, Germany.

Schüürmann G, Ebert RU, Nendza M, Dearden JC, Paschke A, Kühne R (2007). Predicting fate-related physicochemical properties. In: van Leeuwen CJ, Vermeire RG (eds) Risk assessment of chemicals: An introduction 2nd Edition. Springer, Dordrecht, The Netherlands, pp 375-426.

Schüürmann G, Klein W (1988). Advances in bioconcentration prediction. Chemosphere 17(8):1551-1574.

Schultz TW, Diderich R, Kuseva CD, Mekenyan OG (2018). The OECD QSAR Toolbox Starts Its Second Decade. In: Nicolotti, O. (eds) Computational Toxicology. Methods in Molecular Biology, vol 1800. Humana Press, New York, NY, USA. https://doi.org/10.1007/978-1-4939-7899-1_2.

Schwarz MA, Gobas FAPC (2003). A model for estimating bioaccumulation of hydrophobic organic chemicals in birds and mammals. Environ Toxicol Chem 22(4):741–749. https://doi.org/10.1002/etc.5620220402.

Schwarzenbach RP, Gschwend PM, Imboden DM (1993). Environmental organic chemistry. John Wiley & Sons Inc., New York, NY, USA.

Seydel JK, Wiese M (2002). Drug-membrane interactions, analysis, drug distribution, modeling. Methods and Principles in Medicinal Chemistry. Wiley-VCH, Weinheim, Germany.

Sijm DTHM, Verberne ME, Dejonge WJ, Part P, Opperhuizen A (1995). Allometry in the uptake of hydrophobic chemicals determined in vivo and in isolated perfused gills. Toxicol Appl Pharmacol 131, 130-135.

Slater AM (2014). The IUPAC aqueous and non-aqueous experimental pKa data repositories of organic acids and bases. Journal of Computer-Aided Molecular Design 28(10):1031-1034. https://doi.org/10.1016/S0045-6535(01)00054-6.

Smith KE, Rein A, Trapp S, Mayer P, Karlson UG (2012). Dynamic passive dosing for studying the biotransformation of hydrophobic organic chemicals: microbial degradation as an example. Environ Sci Technol 46(9): 4852–4860. https://doi.org/10.1021/es204050u.

Steinmetz FP, Enoch SJ, Madden JC, Nelms MD, Rodriguez-Sanchez N, Rowe PH, Wen Y, Cronin MTD (2014). Methods for assigning confidence to toxicity data with multiple values — Identifying experimental outliers. Science of The Total Environment 482-483:358-365. https://doi.org/10.1016/j.scitotenv.2014.02.115.

Sushko I, Novotarskyi S, Körner R, Pandey AK, Rupp M, Teetz W, Brandmaier S, Abdelaziz A, Prokopenko VV, Tanchuk VY, Todeschini R, Varnek A, Marcou G, Ertl P, Potemkin V, Grishina M, Gasteiger J, Schwab C, Baskin II, Palyulin VA, Radchenko EV, Welsh WJ, Kholodovych V, Chekmarev D, Cherkasov A, Aires-de-Sousa J, Zhang QY, Bender A, Nigsch F, Patiny L, Williams A, Tkachenko V, Tetko IV (2011). Online chemical modeling environment (OCHEM): web platform for data storage, model development and publishing of chemical information. Journal of Computer-Aided Molecular Design 25(6):533-554. https://doi.org/10.1007/s10822-011-9440-2.

Tebes-Stevens C, Patel JM, Koopmans M, Olmstead J, Hilal SH, Pope N, Weber EJ, Wolfe K (2018). Demonstration of a consensus approach for the calculation of physicochemical properties required for environmental fate assessments. Chemosphere 194:94-106. https://doi.org/10.1016/j.chemosphere.2017.11.137.

Trowell JJ, Gobas FAPC, Moore MM, Kennedy CJ (2018). Estimating the Bioconcentration Factors of Hydrophobic Organic Compounds from Biotransformation Rates Using Rainbow Trout Hepatocytes. Archives of environmental contamination and toxicology. 75(2):295–305. https://doi.org/10.1007/s00244-018-0508-z.

UFZ (Helmholtz Zentrum für Umweltforschung) (2021). ChemProp 7.1.1. https://www.ufz.de/ecochem/chemprop.

Ulrich N, Endo S, Brown TN, Watanabe N, Bronner G, Abraham MH, Goss KU (2017). UFZ-LSER database v 3.2. https://www.ufz.de/lserd.

Ulrich N, Goss KU, Ebert A (2021). Exploring the octanol–water partition coefficient dataset using deep learning techniques and data augmentation. Communications Chemistry 4(1):90. https://doi.org/10.1038/s42004-021-00528-9.

UNEP (United Nations Environment Programme) (2001). Stockholm Convention on Persistent Organic Pollutants (POPs). Retrieved from https://www.pops.int.

UNEP (United Nations Environment Programme) (2007). Report of the Persistent Organic Pollutants Review Committee on the work of its third meeting: Addendum: Risk profile on beta hexachlorocyclohexane. UNEP/POPS/POPRC.3/20/Add.9.

US EPA (1996). Product Properties Test Guidelines OPPTS 830.7560 Partition Coefficient (n-Octanol/Water), Generator Column Method. Washington, DC, USA.

US EPA (2012). Estimation Programs Interface Suite™ for Microsoft® Windows, v 4.11. United States Environmental Protection Agency. https://www.epa.gov/tsca-screening-tools/epi-suitetm-estimation-program-interface.

US EPA (2022). Toxicity Estimation Software Tool T.E.S.T. v5.1.2. The United States Environmental Protection Agency's Center for Computational Toxicology and Exposure. https://www.epa.gov/nrmrl/std/qsar/qsar.html.

US EPA (2023). CompTox Chemicals Dashboard. https://comptox.epa.gov/dashboard/.

Veith GD, DeFoe DL, Bergstedt BV (1979). Measuring and Estimating the Bioconcentration Factor of Chemicals in Fish. J Fish Res Board Can 36:1040-1048.

Wania F, Lei YD, Baskaran S, Sangion A (2022). Identifying organic chemicals not subject to bioaccumulation in air-breathing organisms using predicted partitioning and biotransformation properties. Integrated Environmental Assessment and Management 18(5):1297-1312. https://doi.org/10.1002/ieam.4555.

Wassenaar PNH, Verbruggen EMJ, Cieraad E, Peijnenburg WJGM, Vijver MG (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.

Weisbrod AV, Burkhard LP, Arnot J, Mekenyan O, Howard PH, 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, Bioaccumulative, Toxic Substances. Environmental Health Perspectives 115(2):255-261. https://doi.org/10.1289/ehp.9424.

Xu JY, Wang K, Men SH, Yang Y, Zhou Q, Yan ZG (2023). QSAR-QSIIR-based prediction of bioconcentration factor using machine learning and preliminary application. Environment International 177:108003. https://doi.org/10.1016/j.envint.2023.108003.

Zercher M, Coral JA, Nabb D, Powers G, Jones A, Johanning K (2024). A Comparison of In Vitro Metabolic Clearance of Various Regulatory Fish Species Using Hepatic S9 Fractions. Environ Toxicol Chem 43:1390-1405. https://doi.org/10.1002/etc.5866.

Zhang S, Wang Z, Chen J, Xie Q, Zhu M, Han W (2021). Tissue-Specific Accumulation, Biotransformation, and Physiologically Based Toxicokinetic Modeling of Benzotriazole Ultraviolet Stabilizers in Zebrafish (*Danio rerio*). Environ Sci Technol 55(17):11874-11884. https://doi.org/10.1021/acs.est.1c02861.

Zhang J, Huang Y, Pei Y, Wang Y, Li M, Chen H, Liang X, Martyniuk CJ (2023). Biotransformation, metabolic response, and toxicity of UV-234 and UV-326 in larval zebrafish (*Danio rerio*). Environ Int 174:107896. https://doi.org/10.1016/j.envint.2023.107896.

Zhao C, Boriani E, Chana A, Roncaglioni A, Benfenati E (2008). A new hybrid system of QSAR models for predicting bioconcentration factor (BCF). Chemosphere 73:1701-1707.

Zhao L, Montanari F, Heberle H, Schmidt S (2022). Modeling bioconcentration factors in fish with explainable deep learning. Artificial Intelligence in the Life Sciences 2:100047. https://doi.org/10.1016/j.ailsci.2022.100047.

Zhu M, Xiao Z, Zhang T, Lu G (2025). Construction of interpretable ensemble learning models for predicting bioaccumulation parameters of organic chemicals in fish. Journal of Hazardous Materials 482:136606. https://doi.org/10.1016/j.jhazmat.2024.136606.

10 Supplementary Materials

1. supplementary material_log Kow.xlsx
2. supplementary material_log Koa.xlsx
3. supplementary material_pKa.xlsx
4. supplementary material_BCF QSAR.xlsx
5. supplementary material_substance data endpoints.docx
6. supplementary material_analytics.docx
7. supplementary material_HYBIT_raw_data.docx
8. supplementary material_IVIVE_raw_data.docx

A Annex - Evaluation of test systems

As part of this project, the application of the available alternative methods has been critically assessed in order to integrate them in an integrated assessment strategy enabling their future use as part of the regulatory bioaccumulation assessment of substances within the different EU substance orientated legislations.

First, an initial literature search was performed to collect data on bioaccumulation behaviour of substances that has been obtained via alternative testing methods (i.e., HYBIT and in vitro depletion assays). The literature was searched using Google, Google Scholar and the Fraunhofer eLib Service. For fish BCF values, the OECD QSAR Toolbox was used. The main part of the search was performed in a time frame between November 2022 and January 2023. Smaller, substance focused searches were performed throughout the project. Search terms for Google, Google Scholar and the Fraunhofer eLib Service included the following: "*Hyalella azteca* bioconcentration", "*Hyalella azteca* bioaccumulation", "*Hyalella azteca* toxicokinetics", "hepatic depletion in vitro fish", "hepatic biotransformation in vitro fish", "in vitro fish metabolism", "hepatic depletion in vitro fish", "OECD 319", "OECD TG 319", "fish hepatic S9 metabolism". If data for a specific substance was searched for, the substance name, its CAS number or EC number was inserted into the search. CAS numbers were used to find fish BCF values for specific substances in the OECD QSAR Toolbox. If available and/or applicable, OECD validity criteria were applied to the data for evaluation.

We were able to gather this kind of information on approx. 290 different substances. Data for approx. 50 substances could be found that were based on HYBIT studies. In vitro depletion assay data with hepatic S9 fractions of rainbow trout was found for approx. 200 substances, whereas data for approx. 80 substances were found based on assays conducted with rainbow trout hepatocytes. Additional in vitro depletion assay data that were performed using hepatic rat S9 fractions were found for 15 substances and data for rat hepatocyte-based studies were found for 30 substances.

Based on the data collected in the literature research, seven substances were selected to be tested in 16 experiments which are summarised in Table 24.

Table 24: Conducted laboratory experiments with the selected 7 test substances of the project.

Substance	CAS-Nr.	Test system		Status
Pyrene	129-00-0	HYBIT		Completed
β-HCH	319-85-7	HYBIT		Completed
UV-234	70321-86-7	HYBIT		Completed
Diclofenac (ionised)	15307-86-5 acid 15307-79-6 salt	HYBIT		Completed
Diclofenac (partly neutral)	15307-86-5 acid 15307-79-6 salt	HYBIT		Completed
Fluoxetine (ionised)	54910-89-3	HYBIT		Completed

Substance	CAS-Nr.	Test system		Status
Fluoxetine (partly neutral)	54910-89-3	HYBIT		Completed
D4 (Octamethyl-cyclotetrasiloxane)	556-67-2	HYBIT		Failed: Exposure not possible
Pyrene	129-00-0	in vitro (RT-HEP)		Completed
β-HCH	319-85-7	in vitro (RT-HEP)		Completed
UV-234	70321-86-7	in vitro (RT-HEP)		Completed
D4 (Octamethyl-cyclotetrasiloxane)	556-67-2	in vitro (RT-HEP)		Failed: Exposure not possible
Pyrene	129-00-0	in vitro (rat hepatocytes)		Completed
UV-234	70321-86-7	in vitro (rat hepatocytes)		Completed
Chlorpyrifos	2921-88-2	in vitro (rat hepatocytes)		Completed
D4 (Octamethyl-cyclotetrasiloxane)	556-67-2	in vitro (rat hepatocytes)		Failed: Exposure not possible

A.1 HYBIT (OECD TG 321)

A.1.1 Background information

The *Hyalella azteca* bioconcentration test ("HYBIT") is a non-vertebrate test that resembles OECD TG 305. A multi-laboratory ring trial has previously confirmed the potential of the HYBIT to be used as non-vertebrate alternative to bioconcentration studies according to OECD TG 305 (OECD 2012). Three specific compounds (terbutryn, prochloraz, hexachlorobenzene) were selected for their differing properties (hydrophobicity) and applied in HYBITs under semi-static and/or flow-through conditions. The ring trial demonstrated that the HYBIT protocol is robust and transferable among laboratories, allowing for comparable performance even in labs with limited prior experience with *H. azteca* or bioconcentration testing (OECD, 2024a). However, the trial also helped to identify critical aspects of conducting bioconcentration studies. In a few cases, considerable variability in BCF estimates was observed between the participating laboratories. In this context, the need for correct lipid normalisation of the results obtained and the uniform application of the test methods was emphasised in order to compensate for differences and ensure the comparability of the results.

The ring-trial (OECD 2024b) confirmed that OECD TG 321 permits the assessment of the bioconcentration potential of neutral organic substances of different hydrophobicity. In a previous study (Schlechtriem et al. 2019) various substances with differing hydrophobicity, specifically within the range of log K_{OW} values from 2.4 to 7.6, were tested. The BCF values obtained from *Hyalella azteca* tests have shown a strong correlation with BCF in fish for a broader range of hydrophobic compounds (Schlechtriem et al. 2019). The bioconcentration of

hydrophobic organic chemicals is significantly influenced by the lipid content of the organism. Lipid normalisation is thus required to standardise the bioconcentration factor (BCF) values to a common lipid content, allowing for more accurate comparisons between different studies and chemicals. HYBIT BCFs are usually expressed as normalised to a default lipid content of 3%, based on the whole-body wet weight of the amphipods (OECD 2024a). This means that the reported BCF values are adjusted to reflect an amphipod with a lipid content of 3%, thus minimising the impact of naturally occurring variability in lipid levels. However, inadequate extraction can lead to incorrect lipid normalisation, affecting the reliability of BCF estimates. The use of standardised extraction protocols is crucial to achieve consistent lipid measurements across laboratories (OECD 2024a). Differences in lipid measurement techniques and procedures can lead to significant inter-laboratory variability in lipid content readings as demonstrated for OECD TG 305 studies (Schlechtriem et al. 2012). This inconsistency can complicate comparisons and interpretations of BCF values.

In addition, a laboratory-based applied research project commissioned by the German Environment Agency was carried out to refine and broaden the HYBIT bioaccumulation test, demonstrating its applicability to diverse substances and supporting its regulatory acceptance as a non-vertebrate alternative to fish testing (Schlechtriem et al. 2022). The ring-trial primarily validated HYBIT using a set of standard organic chemicals, the project extended testing to include challenging substance classes such as highly hydrophobic UV-stabilisers, ionic organics (e.g., PFAS), and metallic or nanoparticulate material (Ag, TiO_2, Au). It also compared semi-static and flow-through exposure systems and investigated solvent-facilitated versus solvent-free application. These enhancements demonstrated HYBIT's robustness, expanded its applicability beyond the original validation, and strengthened its potential as a non-vertebrate alternative to fish bioaccumulation testing in line with the 3R principles.

Ebert et al. (2023a) discuss several key mechanisms of bioconcentration in the freshwater amphipod *H. azteca*, with a focus on processes particularly relevant to superhydrophobic compounds. One major factor is the strong binding of such substances to organic material, especially total organic carbon (TOC) present in the test medium. This binding significantly reduces the freely dissolved, bioavailable fraction of the compound, thereby affecting uptake rates. This effect is especially pronounced for substances with high log K_{OW} values and has little to no impact on compounds with lower hydrophobicity (Böhm et al. 2016; Burkhard 2000).

In addition to uptake, elimination pathways may differ for highly hydrophobic substances. Conventional elimination mechanisms—such as passive diffusion or metabolic breakdown—may be less relevant, while fecal excretion may play a more dominant role. These complex mechanisms influence not only *H. azteca* tests but also apply to fish, underscoring the need for careful interpretation of bioaccumulation data across species.

Within the HYBIT protocol (OECD TG 321), these issues are addressed by maintaining very low TOC levels in the test system, helping to minimise artefacts related to organic carbon binding. The small-scale design of the HYBIT setup also offers practical advantages, particularly when dealing with superhydrophobic chemicals.

However, for certain superhydrophobic substances, maintaining stable exposure concentrations remains a significant challenge. In these cases, solvent-facilitated application methods (Schlechtriem et al. 2017) may be required, especially in flow-through setups, to ensure consistent exposure conditions throughout the duration of the test, if a stable exposure concentration can be achieved at all.

Testing superhydrophobic compounds

The HYBIT has proven applicable for a broad spectrum of hydrophobic substances and present several practical advantages over traditional fish-based methods. These include the need for smaller test systems, lower substance quantities, and generally reduced costs and resource use (Schlechtriem et al. 2019). The smaller body size and differing physiological characteristics of *H. azteca* often lead to faster uptake and elimination kinetics, allowing steady state to be reached more quickly than in fish. This feature supports the use of *H. azteca* as a potentially efficient and ethically preferable alternative for assessing the bioaccumulation potential of moderately to highly hydrophobic chemicals.

However, when it comes to testing superhydrophobic compounds with a (log K_{OW} > 8, such as certain benzotriazole ultraviolet stabilisers like UV-234, critical limitations arise. Model estimations and experimental studies indicate that achieving steady state may require exposure periods longer than one month, even with *H. azteca* (Ebert et al. 2023a). Such prolonged durations are impractical within the standard protocols of both HYBIT (OECD TG 321) and fish tests (OECD TG 305). Although a kinetic BCF can be estimated from uptake and depuration rate constants, so that reaching steady state is not absolutely necessary, a test duration that is too short in comparison to the time needed to reach steady state or the time needed to reduce the body burden to a certain percentage of the initial concentration (e.g. 95% loss) will lead to incorrect BCF estimations.

The influence of solvent-facilitated application methods on the test system

Solvent-facilitated application methods can cause biofilm build-up in the test vessel during exposure to the test substance and should thus be used with caution as the presence of biofilm in the test vessels can potentially lead to an unacceptable dietary uptake of the test chemical alongside the uptake from the water phase. Since *H. azteca* grazes on, e.g., microbial mats, the contaminated biofilm may likely be ingested by *H. azteca* and thus, a dietary uptake of the substance may likely occur alongside the uptake from the water phase. The observed endpoint would therefore not be a BCF, but rather a BAF. The extent to which the uptake via food impacts the overall uptake may be determined in a separate BMF study as it has been done before with the example of ^{14}C-radiolabeled methoxychlor (Kosfeld et al. 2020).

Testing ionisable organic compounds

So far, there is only limited knowledge on the bioconcentration of IOCs in *H. azteca*. Given the complex interactions between pH, ionisation, and bioaccumulation potential, further studies are recommended to investigate the effects of pH on the bioaccumulation of IOCs in the amphipods.

Testing per- and polyfluoroalkyl substances (PFAS)

PFAS (per- and polyfluoroalkyl substances) are highly persistent synthetic chemicals that bioaccumulate in aquatic organisms mainly through strong protein binding rather than lipid partitioning, leading to accumulation in blood, liver, and kidney, with slow elimination influenced by chain length and functional group (Martin et al. 2003). Bioconcentration studies with *H. azteca* were carried out to assess the bioaccumulation potential of per- and polyfluoroalkyl substances (PFAS) (Schlechtriem et al. 2022). BCF values for PFOS were found to be between 111 and 259 for different test conditions, indicating that while PFOS is bioavailable, it does not accumulate highly in this organism. This correlates with results from studies conducted with other aquatic invertebrate species, suggesting that *H. azteca* can serve as a viable model organism for evaluating PFAS bioaccumulation potential (Schlechtriem et al. 2022). However, in comparison to fish studies (Martin et al. 2003, Gust et al. 2025) the BCF values obtained for PFOS in *H. azteca* are significantly lower than those observed in fish, where kinetic

BCF values can exceed 1000. This discrepancy illustrates the influence of species-specific differences in the bioaccumulation mechanisms of PFAS or of insufficient normalisation of the results obtained. Understanding these interactions is crucial for risk assessment and regulatory decision-making regarding PFOS and related per- and polyfluoroalkyl substances (PFAS). Ongoing research is necessary to fully elucidate the mechanisms and consequences of protein interactions on the bioaccumulation of PFAS. In this context *H. azteca* may provide different insights compared to traditional fish models. But this also means that predictions regarding PFAS bioaccumulation based on *H. azteca* results must be approached cautiously given the different modes of action and interaction in biological systems. Additional investigations into the bioaccumulation of a broader array of PFAS compounds using *H. azteca* are required.

Testing surfactants

Surfactants are amphiphilic molecules with distinct hydrophobic and hydrophilic components. The hydrophilic head of each surfactant is electrically charged. Depending on the charge of the hydrophilic head, the surfactant is classified as anionic, non-ionic, cationic or amphoteric. Lauric acid is a surfactant with emulsifying properties. Raths et al. 2020 investigated the bioconcentration of laurate in *H. azteca,* highlighting the need for more research into how different types of ionisable organic compounds behave in aquatic environments. This will help to develop more comprehensive models for predicting bioaccumulation potential across a wider range of substances. In this context the n-octanol/water distribution coefficient (log D) is of importance which allows not only to assess the partitioning behaviour of chemicals but also considers the concentration of the compound in different ionisation states, which is essential for the bioaccumulation assessment of ionisable surfactants (Hodges et al. 2019).

Additional options during testing

Literature shows that in addition to bioconcentration factors, metabolism rates and metabolite profiles of the test substance can also be determined. Examples are the fungicides azoxystrobin and prochloraz (Fu et al. 2018), the herbicide terbutryn and the fungicide trifloxystrobin (Kosfeld et al. 2020), as well as the pharmaceutical diclofenac (Fu et al. 2020). The only changes to the test system that are necessary to obtain these data sets are additional samplings. The sample sizes for metabolite screenings may have to be increased from 20 to 30 amphipods per sample. Analytical procedures need to be developed to detect and quantify the metabolites. HYBIT experiments can also involve ^{14}C-radiolabel testing which provides additional analytical options in the context of metabolism related questions (*cf*. Schlechtriem et al. 2019, OECD 2024b).

A.1.2 HYBIT (OECD TG 321) experiments: Material and Methods

In this project, the tests were performed under flow-through conditions. Copper reduced, aerated, de-chlorinated and charcoal filtrated tap water was used as dilution water. The substance was provided via basic or stock solutions. Basic solutions are concentrated solutions of the test substance in dilution water. Stock solutions are concentrated solutions of the test substance in a suitable solvent. A peristaltic pump (Ismatec IPC 8) or a syringe pump (Legato 200, KD Scientific) delivered the basic solution or solvent stock into a mixing chamber. A magnetic dosing pump (ProMinent gamma/X) delivered the dilution water to the mixing chamber. The test medium with the desired concentration was thus prepared in the mixing chamber. Via an overflow the prepared test medium entered the test aquarium.

Contrary to the demand by the OECD TG 321, a solvent control was not performed in tests where a solvent was used to maintain stable water concentration of the test substance.

Water samples were taken on a daily basis during the uptake phase. An additional sample was taken shortly after the onset of the depuration phase.

Water quality for the following parameters was determined on a daily basis: pH, temperature, oxygen saturation. Furthermore, the flow-rate was checked daily. Ammonia, nitrite, nitrate and NPOC were determined at the beginning and end of both, uptake and depuration phase.

The duration of uptake and depuration phase were estimated from the test substances' respective log K_{OW} values. Guidance is provided in §28 of OECD TG 321 (OECD 2024a). For ionisable compounds the log D values of the ionised and neutral species of the test substances were used for the respective estimations. In the study of UV-234 the uptake phase was set to 10 days to mimic the uptake phase of the UV-234 study of a former UBA project, which indicated that steady-state was reached (Schlechtriem et al. 2022).

Each *H. azteca* sampling consisted of 3 replicate samples of 20 amphipods each. Additional *H. azteca* triplicate samples were drawn at the beginning of the uptake phase and the end of uptake and depuration phase. These samples consisted of 10 amphipods, each, and were used for lipid determination.

Table 25 summarises the test setup details that were applied to facilitate the HYBIT experiments with the selected test substances.

Table 25: Test setup details for the HYBIT experiments performed as part of this project.

	Pyrene	β-HCH	UV-234	Diclofenac Ionised	Diclofenac Partly neutral	Fluoxetine ionised	Fluoxetine Partly neutral
Uptake duration	6 days	4 days	10 days	2 days	5 days	6 days	6 days
Uptake samplings	8	8	8	9	8	10	10
Depuration duration	6 days	4 days	9 days	2 days	5 days	4 days	4 days
Depuration samplings	7	7	7	7	7	7	6
Lipid samplings	3	3	3	3	3	3	3
Solvent	None	None	DMSO (0.002%)	None	None	None	None
Daily medium exchange	5-fold	5-fold	15-fold	5-fold	5-fold	5-fold	5-fold
Flow rate (total)	34.7 mL/min	34.7 mL/min	104 mL/min	34.7 mL/min	34.7 mL/min	34.7 mL/min	34.7 mL/min
Flow rate solvent	N/A	N/A	2 µL/min	N/A	N/A	N/A	N/A
Flow rate buffer	N/A	N/A	N/A	N/A	0.347 mL/min MES	0.347 mL/min MES	0.347 mL/min MES

	Pyrene	β-HCH	UV-234	Diclofenac Ionised	Diclofenac Partly neutral	Fluoxetine ionised	Fluoxetine Partly neutral
					3.5µL/min HCl	3.5µL/min HCl	8µL/min NaOH
Buffer concentration	N/A	N/A	N/A	N/A	10mM MES	10mM MES	10mM TRIS
Nominal pH	Natural (7.5 - 8.2)	Natural (7.5 - 8.2)	Natural (7.5 - 8.2)	Natural (7.5 - 8.2)	5	6	9
Nominal test concentration	10µg/L	10µg/L	1µg/L	50µg/L	50µg/L	50µg/L	90µg/L
Water samplings	7 Duplicates	7 Duplicates	14 Triplicates	5 Triplicates	7 Triplicates	9 Triplicates	10 Triplicates
Special aspects	None	None	Aquarium was changed after 5 days of uptake	None	None	None	None
Analytical method	GC-MS	GC-MS	LC-MS	LC-MS	LC-MS	LC-MS	LC-MS

The supervision of the analytical work and the calculation of substance concentrations in samples (water, amphipods) collected during the HYBIT experiments were carried out by Dr. Bettina Dudek and Dr. Carolin Mörtl (UV-234, Diclofenac, Fluoxetine), Claire MacKenzie (Pyrene, ß-HCH), and Georg Rademacher (D4), all Fraunhofer IME, Schmallenberg.

A.1.3 Special test design solutions for the UV-234 HYBIT

Due to the high hydrophobicity of the test substance, a preliminary equilibration phase of three days was conducted to saturate all surfaces with the test substance. To prevent loss of the test substance at the surface of the pipette during water sampling, a volumetric pipette (glass) was also equilibrated inside the test vessel. It remained in the test vessel during the complete test. The uptake phase lasted 10 days to mimic the study design in Schlechtriem et al. (2022) for better comparison of the experiments. After 5 days, the exposure vessel was changed to a backup vessel that has also been equilibrated for 3 days prior to insertion of the amphipods. The concentration of the test substance in the water phase had been monitored during the equilibration and the uptake phase on a daily basis in both vessels using triplicate samplings. For the depuration phase, a new aquarium was used and no stock solution was fed into the system. The depuration phase was set to 9 days.

At t = 312, 339 and 384 hrs only 2 instead of three replicates were drawn for each *H. azteca* sampling.

A.1.4 Data evaluation

The concentration data obtained in the HYBIT experiments were evaluated as described in OECD TG 321 Annex 5. BCFs were calculated by sequentially determining k_2 and k_1. BCF errors were calculated via the general law of error propagation using the errors of the rate constants k_1 and k_2 and the SDs of TWA and *H. azteca* tissue concentration as input (Schlechtriem et al. 2019, Mandel 1984).

Preliminary experiments to evaluate exposure settings

For two substances, preliminary experiments were performed: UV-234 and D4. Both substances display elevated hydrophobicity. In addition, D4 is also volatile.

A.1.4.1 UV-234

General analytical remarks

The high hydrophobicity of the test substance led to adsorption to surfaces for all aqueous solutions. Thus, pre-equilibration of all containers, which were in contact with e.g. the test medium, was necessary. Equipment for water sampling (pipettes, measuring cylinder etc.) was equilibrated in the test aquarium itself to prevent losses of the test substance. Water samples were either directly mixed with solvent in the ratio 1:1 or submitted to liquid-liquid extraction.

The addition of an isotope-labelled internal standard was crucial to reduce the variability of measurements.

Evaluation of exposure methods

Solvent-free approaches

Two solvent-free exposure approaches were evaluated and both showed very high variability of the determinable water concentrations. Several reasons for these observations are possible: An inhomogeneous distribution of the test substance in the water phase, precipitates of the test substance, adsorption to particles in the aquaria or matrix effects during analytical measurement. Different strategies were tested to reduce the variability. Among them were centrifugation of the samples to remove particles loaded with test substance and matrix-matched calibrations or standard addition to eliminate matrix effects. Despite all efforts, no stable water concentrations were measured.

Approach 1: Column generated exposure

A column was packed with Florisil loaded with the test substance. The test basin was equilibrated for 40 days with the diluted column eluate and samplings took place at 9 days during this period. No stable concentrations could be determined. Relative standard deviations ranged from 36.4% - 79.4% and mean values (based on five measurements each) ranged from 4.5 – 39.4 µg/L with no apparent trend visible.

Approach 2: Water accommodated fraction

Two different WAF flasks were prepared. One had a nominal UV-234 concentration of 40 µg/L, the other flask was prepared with a nominal concentration of 10 µg/L. Both WAFs were prepared by distributing an appropriate volume of a UV-234 stock solution in MeOH onto the inner walls of a 2 L amber glass flask. After evaporation of the solvent with an N_2 stream, 2 L of dilution water and a stir bar were added to the flask and the flask was placed onto a magnetic stirrer. After one, two, and three days, both flasks were sampled after one hour without stirring. The first 100 mL were discarded, and 5 mL of sample were extracted with heptane, concentrated to dryness, and redissolved in ACN – MeOH 4:1. Relative standard deviations ranged from 23.9 –

147.9% based on five measurements per sampling. Recoveries of nominal concentrations were not consistent.

Solvent-facilitated approaches

During the solvent-facilitated approaches, water samples were processed similar as described for the water accommodated fraction approach (liquid-liquid extraction). For comparison, water samples stabilised with MeOH 1:1 were also measured. The acetonitrile and acetone methods showed differing results. In samples taken during the experimental equilibration phase that used DMSO as solvent, particles were visible after re-dissolution of the dried extract and thus, only direct measurement was performed from that point on.

Solvent-facilitated exposure (acetonitrile, 0.005%)

A solvent-facilitated exposure with 0.01% acetonitrile was used in a former HYBIT study (Schlechtriem et al. 2022) and an increased build-up of biofilm was detected. To avoid or at least reduce this build-up, the content of acetonitrile was halved for this equilibration phase. First samples were drawn after 3 days of equilibration. At this time already, white particles were visible in the water phase. At the following day, a biofilm was visible in the aquarium, and the test was aborted. The variability between replicate water samples ranged from 2.7 – 21.3% (relative standard deviation).

Solvent-facilitated exposure (acetone, 0.002%)

The solvent-facilitated exposure was repeated by changing the acetonitrile stock solution against a stock solution made from acetone. Furthermore, the content of solvent in the exposure water was once again more than halved. The aquarium was monitored for 7 days and triplicate samples were drawn at days 1, 2, 3, 4, and 7. After 3 and 4 days, white particles started to appear and at day 7, a biofilm was visible. The UV-234 concentrations as well as the triplicate measurements (relative standard deviations 2.8 – 90.6%) showed high variability.

Solvent-facilitated exposure (DMSO, 0.002%)

The solvent-facilitated exposure was repeated by changing the acetone stock solution against a stock solution made from DMSO. The aquarium was monitored for 14 days with samplings being performed at days 1, 2, 3, 6, 7, 8, 9, 10, 13, and 14. At days 2 and 3, very high mean concentrations (2.52–3.47 µg/L, nominal concentration 1 µg/L) with variation of triplicates were measured, which indicated that an equilibration phase of the test system was needed. Between day 6 and 14 comparable concentrations between 0.79 and 1.26 µg/L were determined with one lower value at day 10 (0.54 µg/L). The lower concentrations on day 10 as well as on day 1 could be explained by non-saturated equipment. Day 1 was close to test start and on day 9 the aquarium was cleaned. First signs of biofilm formation were observed at day 6. Relative standard deviations of triplicate samples were between 1.3 and 19.9 %.

Due to the acceptable concentrations with DMSO-assisted exposure, this approach was chosen for the HYBIT test for UV-234. Due to biofilm formation after day 6, a change of the aquarium after 8 days was performed.

A.1.4.2 D4

Test system considerations

D4 is a highly hydrophobic and volatile substance that belongs to the class of cyclic volatile methyl siloxanes (cVMS), which can hydrolyse in water. All these aspects should be considered in the test system design for this study.

Instead of a regular, rectangular 20 L aquarium cylindric vessels were chosen (*cf*. Figure 16). The cylindric shape of these vessels provide a smaller water surface which should reduce the chance of substance evaporation from the test medium. In addition, the opening of the cylindric test vessel is smaller than that of the rectangular aquaria, thus it is easier to cover to prevent further substance loss. The cylindric test vessels had already been used in former *H. azteca* bioconcentration tests and have therefore been proven to be appropriate for test conduction (Schlechtriem et al. 2022). The use of silicone materials should be minimised as much as possible to prevent contaminations.

D4 can hydrolyse at pH levels below and above a value of 7 (Brooke et al. 2009). The supplied water in the Fraunhofer IME test facility usually has a pH value of approx. 7.5 – 8, which may already increase the chances of hydrolysis. Therefore, it was tested whether the pH can be adjusted under flow-through conditions. A phosphate buffer was used at a pH of 7. A 100x stock was supplied into the test system and there it was diluted by a factor of 100 via flow-rate adjustment. The final concentrations were 5.2 mM/L KH_2PO_4 and 8.2 mM K_2HPO_4 in the medium inside the test vessel. Determinations of the pH value in the test vessel revealed that the start pH value of 7.76 could be reduced to 7.16 within a day. First changes were seen as soon as 1 hour after start of the buffer addition, when the pH value in the test vessel was already reduced to 7.52. After 4 hours a pH value of approx. 7.2 was reached.

Since the phosphate buffer may interfere with physiological processes of *H. azteca*, it was tested whether the amphipods can survive in the buffered test medium. To evaluate potential toxic effects, 100 *H. azteca* were added to the test vessel and were fed ad libitum with DECOTABs. 4 hours after addition of the amphipods to the test vessel, most *H. azteca* were located at the bottom of the test vessel and did not move much. At the next day all *H. azteca* were dead. Due to the fast onset of the toxic effect, the phosphate in the test medium was reduced by a factor of 1000. Measurements showed that the pH of the test medium does not change with this diluted buffer. Accordingly, the use of a phosphate-buffered medium is apparently not easily applicable for HYBITs.

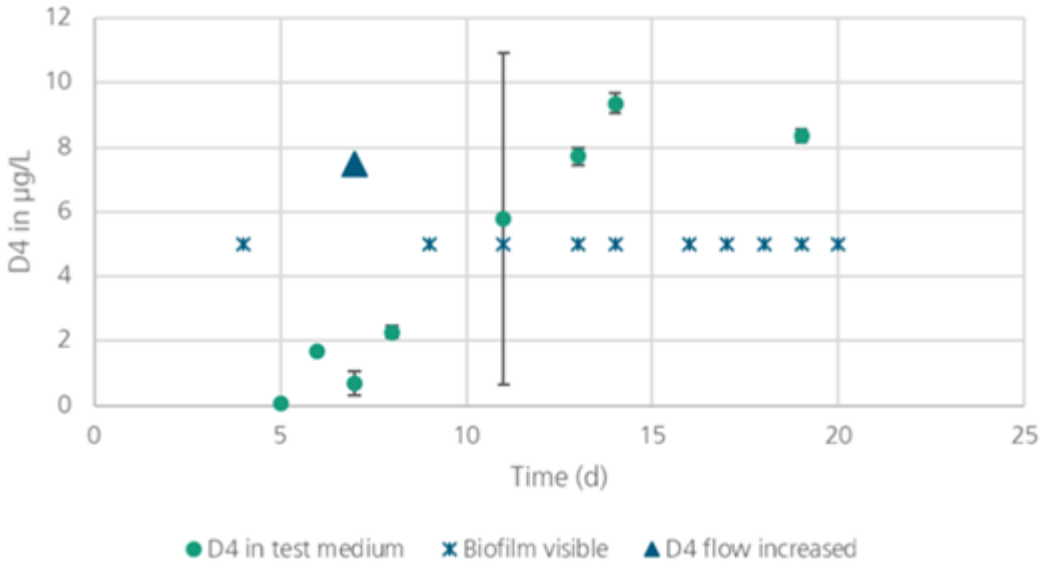
It was also tried to reduce the medium's pH by adding pure HCl to the test medium. However, the HCl damaged the material of the plastic syringe and/or its cannula leading to yellow-greenish colouring of the HCl that is supplied into the test system. Since the HCl was injected at very low rates only (0.97 µL/min), minimum changes in the test system water supply could have a high influence on the resulting pH value. This was reflected in the pH values determined. Using 20% HCl, initial pH values dropped to 4.48 and recovered after 4 hours to a pH value of 7.38. Overnight the pH value reached a value of 8.1, which may show that the 20% HCl may not be sufficient to reduce the pH value permanently. Using 30% HCl, the measured pH value dropped from initially 8.11 to 7.56 after one hour to 5.99 after 3 hours. Using a glass syringe instead of a plastic syringe could reduce the built-up of greenish-yellow colouring in the HCl. In order to maintain relatively stable pH values, the flow rate of the HCl had to be re-adjusted regularly. It was therefore concluded that pure HCl is too error-prone to be used as a pH adjustment in HYBIT flow-through studies.

Preliminary exposure experiments

D4 is a highly hydrophobic substance with a low reported water solubility of approx. 50 µg/L (Brooke et al. 2009). To ensure that the concentration is clearly below the solubility, it was decided that an exposure level of around 10 – 20 µg/L should be chosen. Furthermore, it was decided that the test substance should be administered to the medium using acetone as a solvent carrier (0.0015% v:v). Acetone had been used as a carrier in ^{14}C-D4 BCF studies with fish (Brooke et al. 2009). D4 was introduced into the test system for 22 days and water samples were drawn at 8 days in different compartments of the test system: In the test vessel and at the outlet from the mixing chamber (freshly prepared medium). The application settings were: 10µg/L (nominal) D4 in the test medium, with 0.0015% acetone. A total flow-rate of 2 L/h was applied which results in an approx. 6-fold daily exchange rate.

There were no systematic differences between the concentrations in the mixing chamber and the test vessel. Accordingly, the data evaluation was focussed on the concentration in the test vessel. Results are presented in Figure 14 and Figure 15. Mean concentrations ranged from 0.06 µg/L to 9.4 µg/L with a TWA of approx. 5.6 µg/L. A steady increase of the D4 concentration in the measured test medium samples can be seen. In parallel, biofilm built-up was monitored. Since the water samples were extracted as whole, it cannot be determined whether the measured D4 is present in the water phase or was adhered to co-extracted biofilm particles.

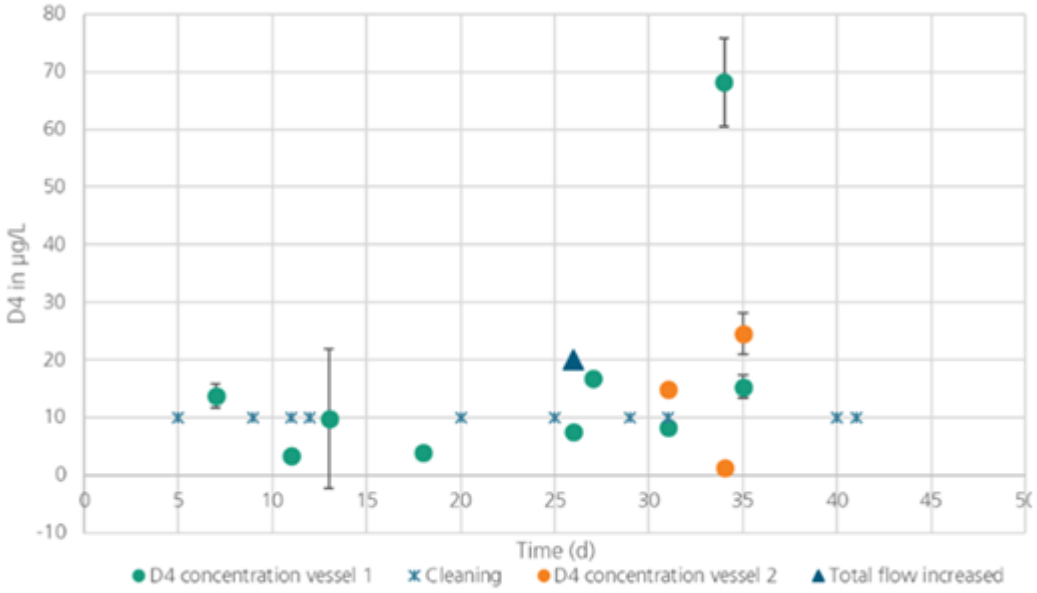
Figure 14: **D4 concentration development in the test medium over time: Fluctuations and biofilm built-up was observed in the test vessel. D4 concentrations are presented as mean values of triplicate measurements, error bars display the standard deviation. At day 5, mean and standard deviation are based on 5 measurements. At day 7 the flow-rate of the D4 stock solution was increased 5-fold.**



Source: own illustration, Fraunhofer IME

The experiment was repeated with the following application settings: A nominal concentration of D4 in the test medium of 50 µg/L, 0.0045% acetone in the test medium and a 6-fold total medium exchange (2 L/h) per day. The inflow was applied for 45 days and at day 26, a second test vessel was installed with the same settings as test vessel 1. At day 21 the settings were changed by increasing the total flow by a factor of 5. Over the course of the test application, samples were drawn at 9 days in vessel 1 and at 3 days in vessel 2.

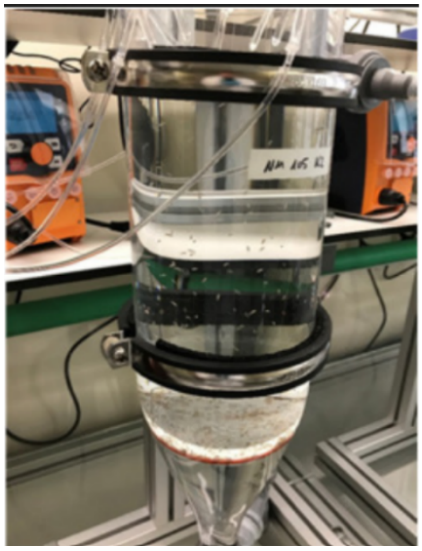
Figure 15: **D4 concentration development in the test medium over time: Fluctuations and biofilm built-up was observed in the test vessel. D4 concentrations are presented as mean values of multiple (3-6) measurements, error bars display the standard deviation. At one time, the total flow-rate of both, the dilution water and the D4 stock solution was increased 3-fold. At the same day, a second test vessel with the same settings as the first one was installed and observed.**



Source: own illustration, Fraunhofer IME

The test application was again marked by biofilm built-up in the test vessels. The test vessels need to be equipped with a sieve at the bottom (*cf.* Figure 16), to prevent the loss of *H. azteca* into the outflow. This sieve was often clogged by the biofilm built-up and at one day the cylindric test vessel overflowed due to severe clogging. To clean the sieve, it needs to be removed from the test vessel. In a test scenario this would require the removal of the *H. azteca* as well. Accordingly, there either has to be a second test vessel with the very same test concentration available with the same, stable concentration of the test substance in the test medium, or the *H. azteca* have to be kept in the water of the test vessel under static conditions as long as the cleaning progresses. The former option was explored by setting up a second test vessel. The latter would require additional experiments to ensure that the D4 concentration remains stable under static conditions despite its volatile nature.

Figure 16: Exemplary picture of the cylindric test vessel set-up for *H. azteca* tests: A sieve has to be installed to prevent *H. azteca* from entering the outflow. Picture published in Kuehr et al. 2021.



The measurements performed in the second application experiment reveal that again no stable concentration could be obtained. Instead of a steady increase, a data scatter was visible. Measured values range from approx. 3.6 – 17.6 µg/L with TWA of 7.9 at a nominal concentration of 50 µg/L. As mentioned above, the test system was again subject to biofilm formation which may have an influence on the measured concentration due to D4 that could have adhered to biofilm particles that were co-extracted.

Finally, it can be concluded that the application of D4 in a stable manner for a HYBIT was not possible. The biofilm formation that was caused by the addition of acetone as solvent may interfere with the measurement of the substance and might explain the data scatter. Further experiments and fine-tuning steps would be necessary to understand whether the biofilm really is a cause of the data scatter and to also identify ways to solve the problem.

Furthermore, the biofilm leads to a clogging of the sieve at the bottom of the vessel. The sieve prevents the *H. azteca* from entering the outflow. A cleaning of the sieve is only possible by removing it which requires a removal of the *H. azteca* as well. Further experiments would be necessary to evaluate techniques that provide adequate accommodation in the meantime at constant concentrations of D4. This was beyond the scope of this project.

Adjustment of the test medium pH

This project also investigated whether the HYBIT test can be performed at different pH levels to investigate the bioconcentration of ionisable substance. A substance is ionisable if, depending on the pH of the surrounding medium, the ionisation of the substance changes. The pH at which the state of ionisation changes is expressed as p*K*a (acids) or p*K*b (bases). If the pH of the surrounding medium falls below an acid's p*K*a value, it loses its charge and becomes increasingly neutral, whereas bases become increasingly neutral if the pH of the surrounding medium reaches a value above their p*K*a. With the state of ionisation, the lipophilicity of the substance may change as well. The log K_{OW} value is not applicable to ionisable substances and thus, the log D is used which expresses the hydrophobicity in dependence with the pH value of the surrounding medium.

It is therefore possible that a substance may exhibit different bioaccumulation characteristics depending on the pH of the surrounding medium. A steady pH value has to be provided in a

HYBIT system to allow for a reliable testing. A set of preliminary experiments was conducted to evaluate techniques, that may be able to provide the required pH stability.

Aeration can lead to a faster drift towards the initial pH of the dilution water, if the pH was only adjusted using an acid (HCl, HNO_3, H_2SO_4) or a base (NaOH) in a static system. Accordingly, it was decided that the test system should be buffered.

In Table 26 all buffers that were evaluated are listed with the respective target pH value. The pH value was stable in a semi-static beaker system. Follow-up experiments exposed a small number of *H. azteca* (3 x 20 *H. azteca*/ treatment) to different concentrations of the buffer substances to evaluate, whether toxicity can be observed. To exclude the possibility that observed toxicity may be due to the buffer substance, *H. azteca* (3 x 20 *H. azteca* per treatment) were also kept in a semi-static, unbuffered system with adjusted pH value. No mortalities that are statistically different from the control (dilution water) with an approx. pH value of 8 were observed for the treatments with TWA pH values of 5.45, 6.35, 7.24, 7.82, 8.67.

Phosphate buffer media (at 10mM, 50mM and 100mM) had white precipitates. The pH value could be maintained at around 10, but it was decided to not use this buffer system for any further testing. One reason is an observed mortality of *H. azteca* at buffer concentrations of 5 – 10mM, the other reason was the presence of the precipitates that may clog the flow-through apparatus. TRIS was not able to keep the pH value stable at a pH value of 10, therefore this buffer system was also not tested at pH 10 with *H. azteca*.

Table 26: Buffer systems tested in preliminary experiments for the pH-adjusted HYBIT system setup.

Buffer	pH 4	pH 5	pH 6	pH 7	pH 9	pH 10
MES	X † (10mM)					
Acetate acetic acid		X				
MES		X	X			
MOPS				X		
TRIS-HCl				X † (50mM)	X	X
Phosphate buffer						X

† = Significant mortality compared to control (dilution water) with corresponding buffer concentration in brackets

To provide a faster pH adjustment in the mixing chamber, 30% HCl/2M NaOH was also added to the dilution water alongside the concentrated buffer stock. With the knowledge gained in the semi-static preliminary tests, the flow-through setup was equipped with the required equipment (additional pumps and mixing chamber) for the application. The pH level of 5, 7, and 10 were selected to represent the minimum and maximum pH values at which no significant mortality of *H. azteca* was observed an in addition, pH 7 as neutral pH was also investigated. Stable pH values were maintained at all evaluated buffer concentrations (10 mM – 50 mM). Care should be taken with the dosing of the HCl, because small, singular droplets can influence the pH value markedly. The test vessel re-equilibrated towards the target pH value after approx. 2 – 3 hours at a total

flow-rate of 2 L/h which corresponds to a 6-fold daily renewal of the test medium in the test vessel (10 L).

The following buffer concentrations were determined adequate for flow-through testing:

pH 5: MES 10 mM

pH 6: MES 10 mM

pH 7: TRIS 10 mM

pH 9: TRIS 10 mM

The 30% HCl was added at a factor 7.9 x 10^6 lower than the total flow-rate for pH 5 and a factor 1 x 10^5 lower than the total flow-rate at pH 6. 2M NaOH was added to the system at a factor 4.3 x 10^6 lower than the total flow rate.

Detailed descriptions of these preliminary tests are available in the Mater Thesis of Kimberly Padberg (Padberg 2024).

A.2 HYBIT (OECD TG 321) experiments: Results

A.2.1 Pyrene

The HYBIT study with pyrene was conducted successfully. Figure 17 shows the results for tissue and water concentrations. The visual inspection of the tissue concentrations shows that steady state has been reached in a time frame ranging from 48 hours to 144 hours (end of uptake). Calculations confirm that the tissue concentrations lie within a ± 20% range around the mean tissue concentration for that time, which confirms the steady state observation.

The water concentration dropped slightly within the first 48 hours of the experiment and remained stable afterwards until the end of the uptake phase. Despite elevated values at the beginning of the study, all mean water concentration data lie within the ± 20% TWA range. The determined TWA of 0.98 µg/L is lower than the reported solubility of 134 µg/L (ECHA registration dossier for EC number: 204-927-3).

The mean lipid value determined for *H. azteca* during the pyrene study was 2.2%. The results of one replicate that was taken at the begin of the experiment had to be excluded from the evaluation since the calculated lipid content was at 5.3% which is considerably higher than all remaining results and thus, this replicate was treated as an outlier.

The calculated mortality during the test was 14.5% and thus below the threshold of 20%. Temperature was held steadily around a mean of 22.5°C and differed by max. ± 1.3°C which is below the allowed ± 2°C. The mean pH value was at 7.68 and no single value crossed the range of 6.5 – 8.5. Oxygen saturations were at a mean of 96.12% (≙ 7.97 mg/L) and did not fall lower than 94.6% (≙ 7.79 mg/L). Total flow rates were kept at steady 2.23 L/h during the uptake phase and 1.68 L/h during the depuration phase, where the total flow rate was reduced. Ammonia and nitrite were non-detectable, nitrate values ranged from 8-9 mg/L and were thus below any threshold of concern. NPOC levels ranged from 0.52 – 0.98 mg/L, which is below the threshold of 2 mg/L (OECD 2024).

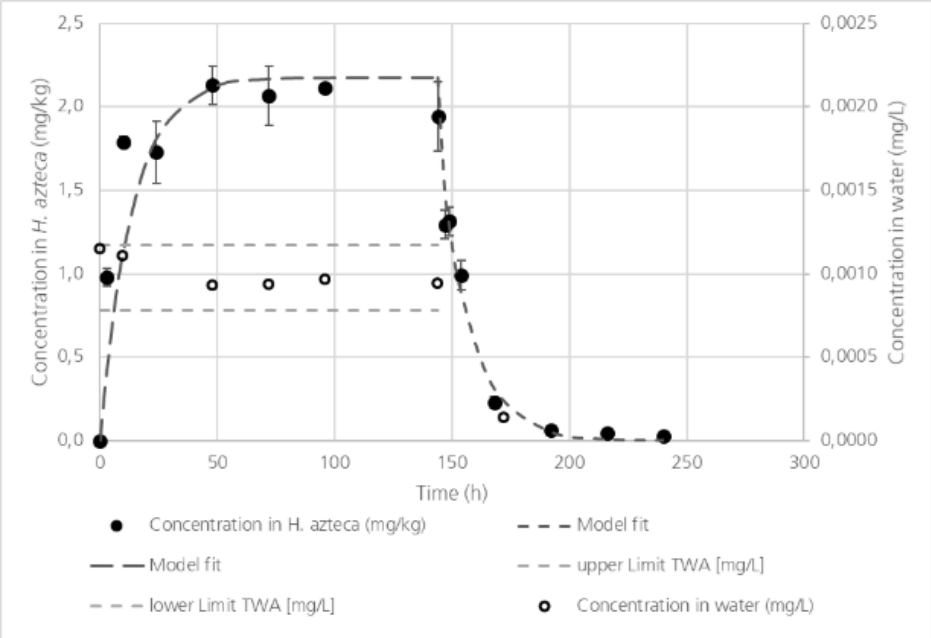
The calculated, non-lipid-normalised BCF values are 2109 ± 282.3 (steady state) and 2225 ± 222.5 (kinetic). The similarity between steady state and kinetic BCF supports the observation that steady state was reached. Table 27 shows the lipid normalised BCF values.

H. azteca bioconcentration tests with pyrene have been performed in the past and literature data is available. However, only radiolabeled ^{14}C-pyrene had been used (Schlechtriem et al. 2019, Lee at al. 2002). Thus, only BCF data that also includes metabolites is available.

Note:

Raw data related to the HYBIT experiments are available as supplementary materials "HYBIT raw data" and "analytics".

Figure 17: Concentration data of the HYBIT study with pyrene. Water concentration data is displayed on a factor 100 lower axis than the *H. azteca* concentration. The displayed model fit was calculated from TWA value, k_1, and k_2 values that are given in Table 27.



Source: own illustration, Fraunhofer IME

Table 27: ^ BCF parameters for the HYBIT study with pyrene.

Parameter	Value	Unit
k_1	165.41 3969.84	L/(kg*h) L/(kg*d)
k_2	0.0743 1.7832	1/h 1/d
$C_{H,SS}$	2.06	mg/kg
TWA	0.98	µg/L
Lipid	2.19 ± 0.73	%
BCF_{SS}	2109 ± 282.3	L/kg
BCF_K	2225 ± 222.5	L/kg

Parameter	Value	Unit
BCF_{SSL}	2891 ± 1036 (3% total lipid) 4818 ± 1727 (5% total lipid)	L/kg
BCF_{KL}	3050 ± 1059 (3% total lipid) 5083 ± 1765 (5% total lipid)	L/kg
Mortality	14.5	%

A.2.2 β-HCH

The HYBIT study with β-HCH was conducted successfully. Figure 18 shows the results for tissue and water concentrations. The visual inspection of the tissue concentrations indicates that steady state may have been reached in a time frame ranging from 48 hours to 96 hours (end of uptake). Calculations confirm that the tissue concentrations lie within a ± 20% range around the mean tissue concentration for that time, which could be an indicator for a reached steady-state. One tissue sample at t = 48 hrs showed exceptionally higher β-HCH concentrations than all other tissue samples and was thus omitted from the data.

The water concentration displays a slow increasing trend throughout the uptake phase. Despite this trend, the mean water concentration of all sampling points lies within the ± 20% TWA range. The determined TWA of 7.65 µg/L is lower than the reported solubility of 200 µg/L (HSDB database, entry 6182, https://pubchem.ncbi.nlm.nih.gov/source/hsdb/6183, last checked January 6th, 2025).

The mean lipid value determined for *H. azteca* during the β-HCH study was 1.7%.

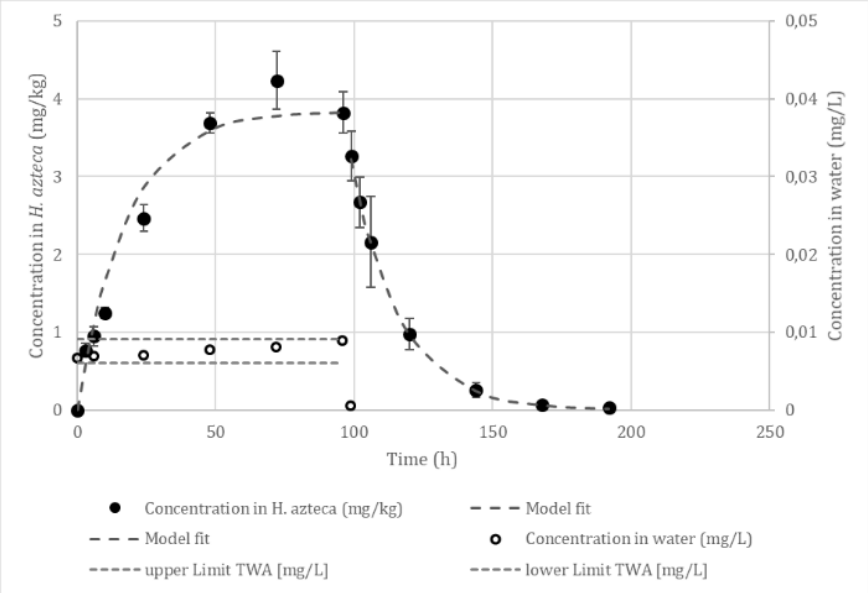
The calculated mortality during the test was 8.0% and thus below the threshold of 20%. The temperature was held steadily around a mean of 23.6°C and differed by max. ± 1.0°C instead of the ± 2°C that would be permissible according to the draft test guideline. The mean pH value was at 7.78 and no single value crossed the range of 6.5 – 8.5. Oxygen saturations were at a mean of 97.02% (≙ 7.93 mg/L) and did not fall lower than 94.6% (≙ 7.77 mg/L). Total flow rates were kept at steady 2.04 L/h for the entire test. Ammonia had a maximum value of 0.06 mg/L but was non-detectable in all other measurements. Nitrite was detected at a concentration of 0.009 mg/L once but was non-detectable in all other measurements. Nitrate values ranged from 5-7 mg/L. None of these values crossed any level of concern. NPOC levels ranged from 0.46 – 2.40 mg/L. The maximum allowed value of 2mg/L was crossed at t = 0 minutes but remained below the threshold for the remaining test.

The calculated, non-lipid-normalised BCF values are 516 ± 73.8 (steady state) and 503 ± 42.9 (kinetic). The similarity between steady state and kinetic BCF furthermore indicate that steady state was reached. Table 28 shows the lipid normalised BCF values.

Note:

Raw data related to the HYBIT experiments are available as supplementary materials “HYBIT raw data” and “analytics”.

Figure 18: **Concentration data of the HYBIT study with β-HCH. Water concentration data is displayed on a factor 100 lower axis than the *H. azteca* concentration. The displayed model fit was calculated from TWA value, k_1, and k_2 values that are given in Table 28.**



Source: own illustration, Fraunhofer IME

Table 28: **BCF parameters for the HYBIT study with β-HCH.**

Parameter	Value	Unit
k_1	28.73 689.52	L/(kg*h) L/(kg*d)
k_2	0.0571 1.3704	1/h 1/d
$C_{H,SS}$	3.94	mg/kg
TWA	7.65	µg/L
Lipid	1.70 ± 0.22	%
BCF_{SS}	516 ± 73.8	L/kg
BCF_K	503 ± 42.9	L/kg
BCF_{SSL}	910 ± 176 (3% lipid) 1516 ± 293 (5% lipid)	L/kg
BCF_{KL}	887 ± 138 (3% lipid) 1478 ± 230 (5% lipid)	L/kg
Mortality	8.0	%

A.2.3 UV-234

The HYBIT study with UV-234 was concluded but did not meet all validity criteria of OECD TG 321. Figure 19 shows the results for tissue and water concentrations.

The visual inspection of the tissue concentrations indicates, steady state has not been reached and accordingly, only the *H. azteca* tissue concentration at the end of the uptake phase was used to calculate a BCF_{SS}. Overall, a gradual increase of the tissue concentration can be seen, followed by a rapid decrease with the onset of the depuration phase. Larger standard deviations are present at time points t = 144 hrs and t = 240 hrs. In addition, at t = 144 hrs the course of concentration appears to fall beneath the expected trend. It is possible that this drop is an artefact that was caused by the vessel change. Accordingly, the tissue concentration data point at t = 144 hrs was excluded from the BCF calculations. At t = 288 hrs one replicate displayed exceptionally high tissue concentrations, which were higher than all other tissue concentrations determined in this study. This replicate was thus excluded from the calculations.

UV-234 water concentrations in the test vessels underwent some fluctuations as shown in detail in Figure 20. The vessel change was performed after t = 120 hrs. The measured water concentration at t = 130 hrs prominently lies above the + 20% TWA value. Excluding this value from the TWA calculation changes the obtained TWA value from 1.00 µg/L to 0.97 µg/L. Since there was no analytically justified reason to question the validity of this data point it was included in the TWA calculation. There is only very limited information available regarding the water solubility of UV-234. The ECHA registration dossier reports a value of < 5 µg/L at 20°C and a pH of 6.1 (ECHA registration dossier for EC number: 274-570-6). The determined values are lower than 5 µg/L.

Due to a malfunction of the lighting control, the 16:8 light:dark rhythm was not maintained during the uptake phase. Instead, the lighting persisted for 24 hrs per day. During the depuration phase the lighting control maintained the 16:8 light:dark rhythm.

The lipid values at the start of the study are considerably lower than the lipid values measured in the middle and at the end of the study. Since the data that was determined in the middle and at the end of the study better reflects the lipid content of the amphipods during the study, the lipid content at t= 0 hrs was not used to calculate the mean lipid content.

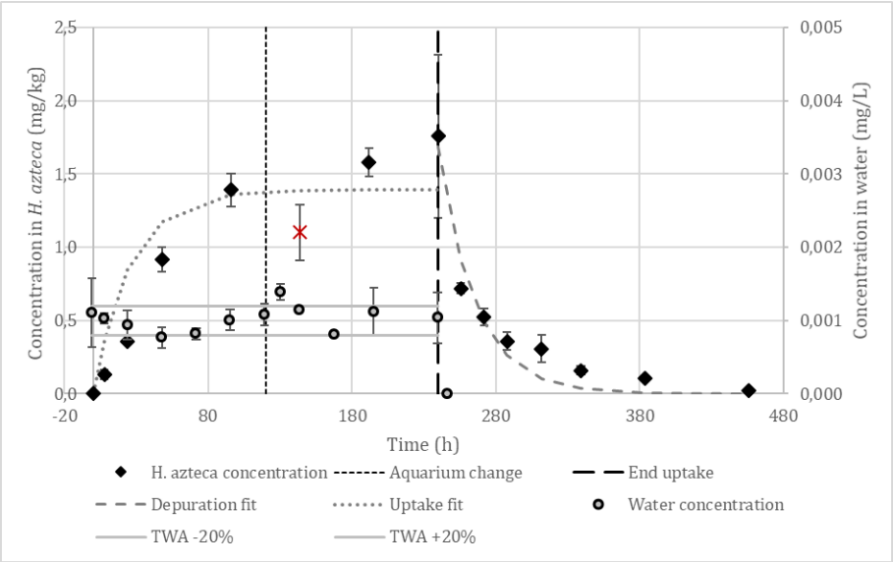
The basic parameters of the HYBIT study for UV-234 were within the acceptable range for a bioconcentration study with *H. azteca*. The water temperature was constant at a mean value of 23.3°C with minimum and maximum values of 22.3 and 23.5°C, respectively. Oxygen saturation could be held at a constant level with values that ranged from 82.9% (6.66 mg/L) – 120.8% (10.76 mg/L) with a mean of 100.2% (8.36 mg/L). The pH value was steadily at a mean value of 8.2 with minimum and maximum values of 8.0 and 8.3. One single measurement gave a pH of 8.6. This is higher than the validity criterium for maximum pH. A mean total flow rate of 6.6 L/h was provided which is equal to a 15.8-fold daily exchange of the exposure medium. A minimum value of 6.1 L/h and a maximum value of 7.2 L/h were determined. Ammonium, nitrite and nitrate levels were consistently low. Ammonium levels ranged from not detectable (< 0.01 mg/L) to 0.5 mg/L. Nitrite had levels of not detectable (< 0.005 mg/L) to 0.005 mg/L. Nitrate was determined at levels of 7 – 15 mg/L. None of these concentrations give rise for concern. NPOC values were higher than in other HYBIT studies which can be explained by the use of DMSO as solvent. NPOC values ranged from 5.5 – 7.5 mg/L during the uptake phase and dropped to 0.7 – 1.1 mg/L during the depuration phase. Approximately 6.4 mg carbon/L can be explained by the presence of 0.0019% DMSO in the test medium. Despite the comparatively long test and two vessel changes, a mortality rate of 18.5%, which is below the 20% threshold, could be determined.

The calculated, non-lipid-normalised BCF values are 1758 ± 699.3 (steady state) and 1394 ± 237.8 (kinetic). The steady state BCF is elevated compared to the kinetic BCF but the kinetic BCF lies within the error margin of the steady state BCF. The relative error of the steady-state BCF is approx. 40%, which could be an important factor when comparing both BCF values. Table 3 shows the lipid normalised BCF values.

Note:

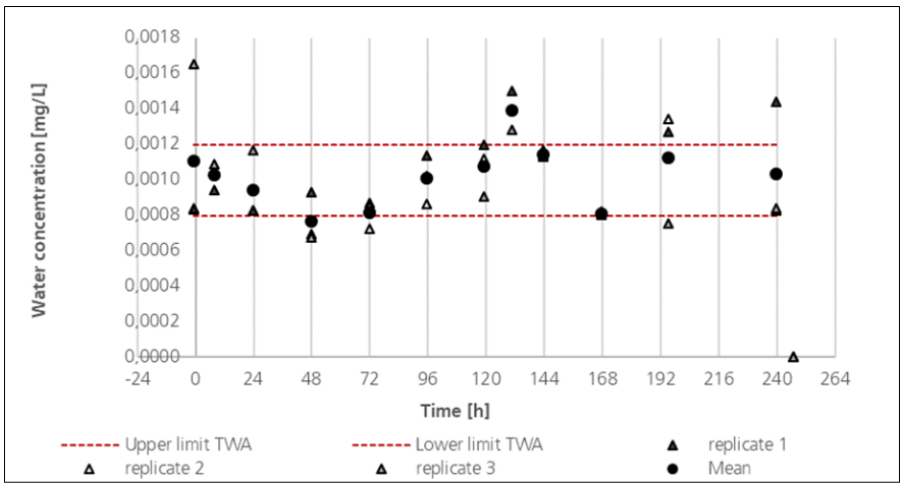
Raw data related to the HYBIT experiments are available as supplementary materials "HYBIT raw data" and "analytics".

Figure 19: Concentration data of the UV-234 HYBIT and the fit results of the tissue concentration data. Marked in red is the tissue concentration data at t = 144hrs which was excluded from the data fit. Water concentration data is displayed on a factor 500 lower axis than the *H. azteca* concentration. The displayed model fit was calculated from TWA value, k_1, and k_2 values that are given Table 29.



Source: own illustration, Fraunhofer IME

Figure 20: Detailed plot of the water concentrations of UV-234 during the HYBIT study. The aquarium change was performed at t = 120 hrs.



Source: own illustration, Fraunhofer IME

Table 29: BCF parameters for the HYBIT study with UV-234.

Parameter	Value	Unit
k_1	53.952 1294.8	L/(kg*h) L/(kg*d)
k_2	0.0387 0.9288	1/h 1/d
$C_{H,SS}$	1.37	mg/kg
TWA	1	µg/L
Lipid	2.5	%
BCF_{SS}	1758 ± 699.3	L/kg
BCF_K	1394 ± 237.8	L/kg
BCF_{SSL}	2106 ± 1019 (3% total lipid) 3510 ± 1698 (5% total lipid)	L/kg
BCF_{KL}	1670 ± 541 (3% total lipid) 2784 ± 902 (5% total lipid)	L/kg
Mortality	18.5	%

A.2.4 Diclofenac (ionised)

Ionizable organic compounds (IOCs) can be either cations or anions, and their charge influences their environmental behaviour and bioaccumulation. The ionisation of IOCs is pH-dependent, with many compounds exhibiting different speciation patterns at various pH levels. This changes their solubility, leading to differences in how readily they can be taken up by aquatic organisms. Generally, the neutral form of a compound is more likely to penetrate biological membranes than its ionised form (e.g. Köhler et al. 2023). The reduced permeability of IOCs can result in lower bioaccumulation potential at certain pH levels where a greater proportion of the compound is in the ionised state. Therefore, experimental assessment of bioaccumulation needs to consider the pH of the test environment, as it could greatly affect the outcome of bioaccumulation studies. Understanding this relationship is vital for accurate risk assessments and predictions regarding the bioaccumulation behaviour of IOCs.

Diclofenac was selected to test, since it is an ionizable acid. Preliminary experiments demonstrated that the in-house *H. azteca* do not survive a pH value of 4, but of 5 (*cf.* chapter 5.2.1.3 “Adjustment of the test medium pH”). With the following values that characterise its ionisable nature (Köhler et al. 2023) it can be concluded that it is possible to test this substance at its ionised and also at its nearly neutral state:

p*K*a: 3.99, 4 00, 4.15

log D (ionic): 0.7

log D (neutral): 4.3

The consensus log K_{OW} determined in this study is 4.34 for diclofenac and 3.09 for the sodium salt of diclofenac.

This chapter describes the results of the test that was performed at the "natural" pH of the holding and dilution water of the test facility, which is approx. 7.5 – 8.2. At this pH value diclofenac is present in its ionised state.

The HYBIT study with Diclofenac (ionised) was conducted successfully. Figure 21 shows the results for tissue and water concentrations. The visual inspection of the tissue concentrations shows that steady state has been reached in a time frame ranging from 23 hours to 48 hours (end of uptake). Calculations confirm that the tissue concentrations lie within a ± 20% range around the mean tissue concentration for that time, which confirms the steady state observation.

The water concentration remained stable throughout the uptake phase. All mean water concentration data lie within the ± 20% TWA range. The determined TWA of 52.78 µg/L is lower than reported solubility levels of diclofenac in water. Ionic strength and the pH of the water have an impact on the solubility of diclofenac (e.g., Kroll et al. 2024), but most reported data for pH levels of 7 - 8 lie in mg/L ranges (see also ECHA registration dossier for EC number 239-346-4).

The mean lipid value determined for *H. azteca* during the diclofenac (ionised) study was 2.07%.

The calculated mortality during the test was 15.9% and was thus below the validity threshold of 20%. Temperature was held steadily around a mean of 23.6°C and differed by max. ± 0.3°C instead of the allowed ± 2°C. The mean pH value was at 7.88 and no single value crossed the range of 6.5 – 8.5. Oxygen saturations were at a mean of 94.92% (≙ 7.70 mg/L) and did not fall lower than 89.3% (≙ 7.31 mg/L). Total flow rates were kept at steady 2.06 L/h during the test. Ammonia was non-detectable. Nitrite values ranged from 0.006 – 0.056 mg/L and nitrate values were steady at 10 mg/L and were thus below any threshold of concern.

Mean NPOC values of 1.77 ± 0.66 mg/L were measured during the test. The measurement that was taken directly after the onset of the depuration phase crossed the limit of max. 2 mg/L that is demanded in OECD TG 321 with a value of 2.65 mg/L. Since this was a singular event and no substance that may bind to TOC was present in the aquarium, this deviation should not have any considerable impact.

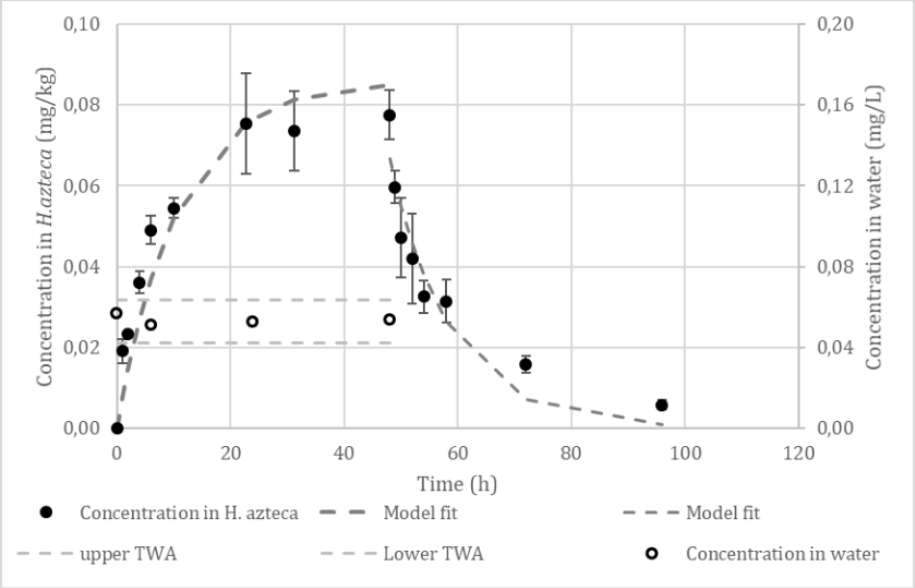
The calculated, non-lipid-normalised BCF values are 1.4 ± 0.2 (steady state) and 1.6 ± 0.3 (kinetic). The similarity between steady state and kinetic BCF supports the observation that steady state was reached. Table 4 shows the lipid normalised BCF values.

The calculated BCF values are very comparable to the kinetic BCF calculated in Kosfeld et al. (2019) which are 1.4 (BCF_k) and 3 (BCF_{kL}, 5% lipid).

Note:

Raw data related to the HYBIT experiments are available as supplementary materials "HYBIT raw data" and "analytics".

Figure 21: **Concentration data of the HYBIT study with diclofenac (ionised). Water concentration data is displayed on a factor 2 higher axis than the *H. azteca* concentration. The displayed model fit was calculated from TWA value, k_1, and k_2 values that are given in Table 30.**



Source: own illustration, Fraunhofer IME

Table 30: **BCF parameters for the HYBIT study with diclofenac (ionised).**

Parameter	Value	Unit
k_1	0.015132 ± 0.00532	L/(kg*h) L/(kg*d)
k_2	0.09292 ± 0.01461	1/h 1/d
$C_{H,SS}$	0.07549 ± 0.01	mg/kg
TWA	0.05278 ± 0.00232	µg/L
Lipid	2.07 ± 0.25	%
BCF_{SS}	1.4 ± 0.2	L/kg
BCF_K	1.6 ± 0.3	L/kg
BCF_{SSL}	2.1 ± 0.4 (3% lipid) 3.5 ± 0.6 (5% lipid)	L/kg
BCF_{KL}	2.4 ± 0.5 (3% lipid) 3.9 ± 0.6 (5% lipid)	L/kg
Mortality	15.9	%

A.2.5 **Diclofenac (partly neutral)**

This chapter describes the results of the diclofenac HYBIT that was performed at a nominal pH of 5. This pH was facilitated by adding HCl and MES buffer (final concentration 10 mM) to the holding and dilution water of the test facility, which has a natural pH of approx. 7.5 – 8.2.

Due to a malfunctioning of one pump after approx. 24 hrs of application that caused drop in the pH value of the medium and also a mortality event during the uptake phase, only two replicates were drawn for each *H. azteca* sampling during the depuration phase. The final lipid sampling at the end of the depuration phase was also reduced to two replicates. Figure 22 shows the drop in the pH value and the consecutive rise in the *H. azteca* tissue concentration. The water concentration of diclofenac remained in a stable ± 20% range around the TWA and was not affected by the drop of the pH value (*cf*. Figure 22). The tissue concentration at t = 48 hrs was omitted from the BCF calculations for the following two reasons: At a pH of 5 not all diclofenac molecules will be present in their neutral form, at pH 3, however, this is more likely since diclofenac has a p*K*a of approx. 4. It is therefore possible that more diclofenac may be retained in the *H. azteca* tissue than at a pH value of 5. During this time an elevated toxicity was observed. This could furthermore increase the accumulation of diclofenac, because the amphipods may change their metabolism to survive the low pH value.

Excluding the tissue concentration data at t = 48 hrs shows that steady state has likely been reached after 24-48 hrs already. Due to the mortality event at t = 48 hrs, only the concentration data at t = 72, 96 and 120 hrs were used to calculate the steady state BCF. Although the tissue concentration data at t = 120 hrs shows a decline, calculations confirm that the mean tissue concentrations at all three sampling times lie within a ± 20% range around the mean tissue concentration, which confirms the steady state observation.

As Figure 23 shows, the mean water concentration remains in a ± 20% TWA range. The determined TWA of 44.85 µg/L lie below reported water solubility values for diclofenac (*cf*. chapter “Diclofenac (ionised)”).

The mean lipid value determined for *H. azteca* during the diclofenac (partly neutral) study was 1.55%.

The calculated mortality during the test was 36.2% and thus higher than the threshold of 20%. An explanation for this increased mortality can be found in the pH drop during the uptake phase and can be assigned to this single event. Temperature was held steadily around a mean of 23.6°C and differed by max. ± 0.3°C instead of the allowed ± 2°C. The mean pH value was at 5.3 with min-max values of 4.88 – 5.58, respectively, if the singular event of pH drop is taken out of the calculation. Oxygen saturations were at a mean of 94.87% (≙ 7.77 mg/L) and did not fall lower than 82.0% (≙ 6.58 mg/L). Total flow rates were kept at steady 2.08 L/h in the test. Ammonia was detected with 0.01 – 0.04 mg/L during the uptake phase and was non-detectable in the depuration phase. Nitrite values ranged from 0.005 – 0.008 mg/L during the uptake phase and was non-detectable in the depuration phase. Nitrate values ranged from 9 - 11 mg/L and were thus below any threshold of concern. With values of 641 – 799 mg/L, NPOC levels were higher than it is allowed in the OECD TG 321. These extraordinary high levels can be explained by the addition of the 10 mM MES as buffer substance, as this accounts for 720 mg/L additional carbon in the medium.

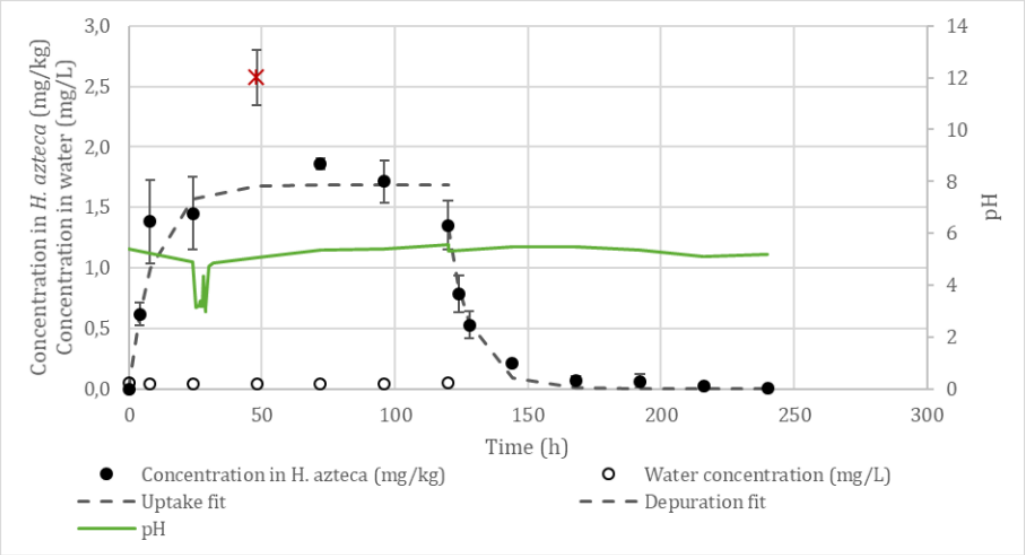
The calculated, non-lipid-normalised BCF values are 36 ± 6.5 (steady state) and 38 ± 5.2 (kinetic). The similarity between steady state and kinetic BCF supports the observation that steady state was reached. Table 31 shows the lipid normalised BCF values.

The change in the pH of the surrounding medium from approx. 8 to approx. 5 increased the bioaccumulation potential of diclofenac by a factor of approx. 27.

Note:

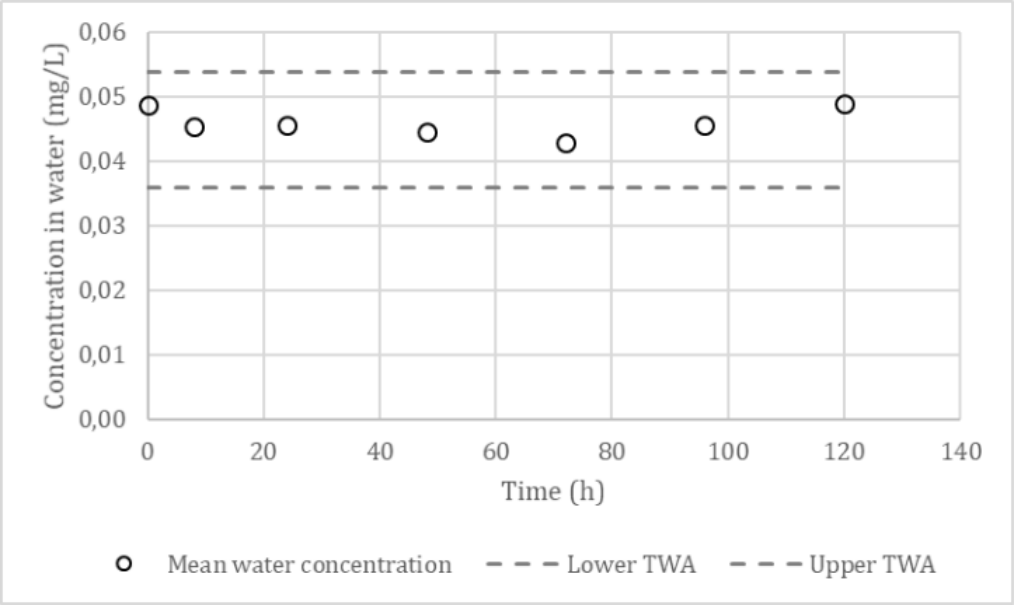
Raw data related to the HYBIT experiments are available as supplementary materials "HYBIT raw data" and analytics".

Figure 22: Concentration data of the HYBIT study with diclofenac (partly neutral). The pH value is presented on the right x-axis. The displayed model fit was calculated from TWA value, k_1, and k_2 values that are given in Table 31. Water concentration is additionally presented in Figure 23.



Source: own illustration, Fraunhofer IME

Figure 23: Detailed plot of the mean water concentrations of diclofenac during the HYBIT study at a nominal pH of 5. The concentration remained stable in a ± 20% TWA range.



Source: own illustration, Fraunhofer IME

Table 31: BCF parameters for the HYBIT study with diclofenac (partly neutral).

Parameter	Value	Unit
k_1	4.1738 ± 0.18863	L/(kg*h) L/(kg*d)
k_2	0.11114 ± 0.0146	1/h 1/d
$C_{H,SS}$	1.594 ± 0.27	mg/kg
TWA	0.04485 ± 0.003	µg/L
Lipid	1.55 ± 0.38	%
BCF_{SS}	36 ± 6.5	L/kg
BCF_K	38 ± 5.2	L/kg
BCF_{SSL}	69 ± 21.1 (3% lipid) 115 ± 35.1 (5% lipid)	L/kg
BCF_{KL}	73 ± 20.5 (3% lipid) 121 ± 34.1 (5% lipid)	L/kg
Mortality	36.9	%

A.2.6 Fluoxetine (ionised)

Fluoxetine was selected to test, since it is an ionisable base. With the following values that characterise its ionisable nature (Köhler et al. 2023) it can be concluded that it is possible to test this substance at its ionised and also at its nearly neutral state:

p*K*a: 9.80, 10.06

log D (ionic): 0.26

log D (neutral): 3.67

The consensus log K_{OW} determined in this study is 4.34 for fluoxetine, 3.99 for the S-enantiomer of fluoxetine, and 4.35 for the R-enantiomer of fluoxetine.

This chapter describes the results of the test that was performed at a nominal pH value of 6. This pH was facilitated by adding HCl and MES buffer (final concentration 10 mM) to the holding and dilution water of the test facility, which has a natural pH of approx. 7.5 – 8.2. At this pH value fluoxetine is present in its ionised state.

The HYBIT study with fluoxetine (ionised) was conducted successfully. Figure 24 shows the results for tissue and water concentrations and the determined pH values. The visual inspection of the tissue concentrations shows that steady state may have been reached in a time frame ranging from 24 hrs to 144 hrs (end of uptake). Calculations confirm that the tissue concentrations lie within a ± 20% range around the mean tissue concentration for that time, which supports the steady state observation. A slight decline in tissue concentration can be

observed, which may indicate that *H. azteca* could have started to adapt to the substance and to increasingly metabolise it.

The water concentration is displayed in more detail in Figure 25. Apart from the measurement at t = 96 hrs, which fell below the -20% TWA limit, all mean water concentrations lie in the ± 20% TWA range. The determined TWA of 49.07 µg/L is lower than reported solubility values that all lie in a mg/L range (e.g., Oakes et al. 2010, Köhler et al. 2023, Nentwig et al. 2007). As reported for diclofenac, ionic strength and the pH of the water will have an impact on the solubility of fluoxetine as well.

The mean lipid value determined for *H. azteca* during the fluoxetine (ionised) study was 3.05%.

The calculated mortality during the test was 14.0% and thus below the threshold of 20%. Temperature was held steadily around a mean of 23.5°C and differed by max. ± 0.1°C which is below the allowed ± 2°C. The mean pH value was at 6.02 with minimum and maximum pH values of 5.92 and 6.13. Oxygen saturations were at a mean of 94.9% (≙ 7.77 mg/L) and did not fall lower than 82.0% (≙ 6.58 mg/L). Total flow rates were kept at steady 2.04 L/h during the test. Ammonia and nitrite were non- to barely detectable; nitrate values were steady at 9 mg/L and were thus below any threshold of concern. NPOC levels ranged from 578 – 856 mg/L, which is much higher than the threshold of 2 mg/L. These extraordinary high levels can be explained by the addition of the 10 mM MES as buffer substance, as this accounts for 720 mg/L additional carbon in the medium.

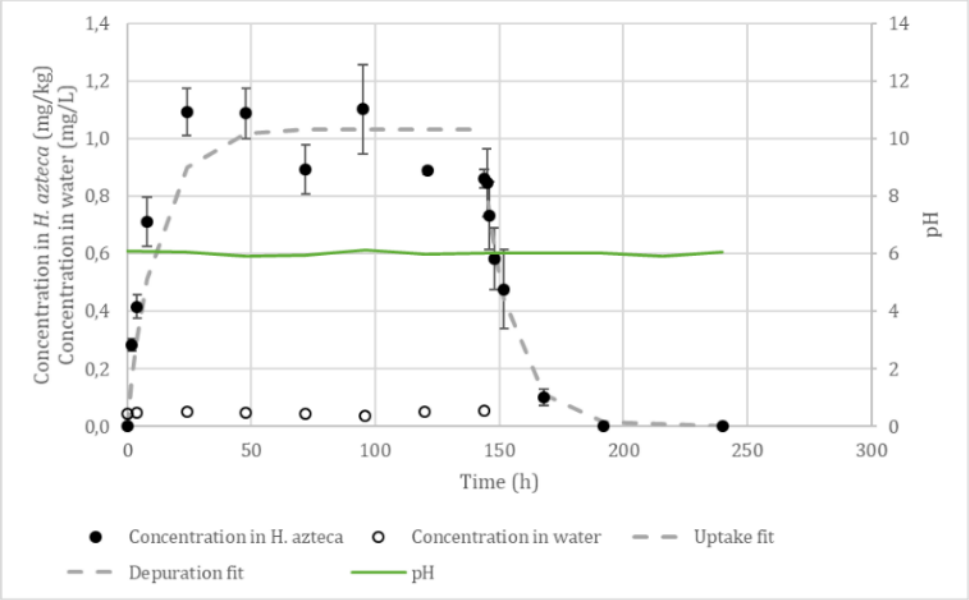
The calculated, non-lipid-normalised BCF values are 20 ± 3.8 (steady state) and 21 ± 2.7 (kinetic). The similarity between steady state and kinetic BCF supports the observation that steady state was reached. Table 6 shows the lipid normalised BCF values.

The metabolite norfluoxetine was also measured in the *H. azteca* samples. The concentration of this metabolite increases constantly and does not reach a steady-state. At the end of the uptake phase the concentration of the metabolite is approximately twice as high as the concentration of the parent substance fluoxetine.

Note:

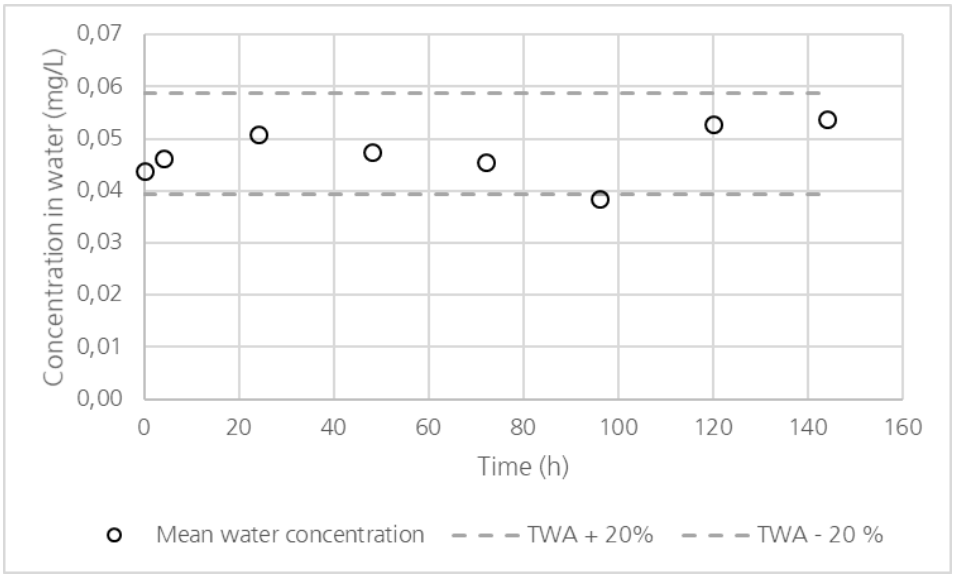
Raw data related to the HYBIT experiments are available as supplementary materials “HYBIT raw data” and “analytics”.

Figure 24: **Concentration data of the HYBIT study with fluoxetine (ionised). The pH value is presented on the right x-axis. The displayed model fit was calculated from TWA value, k_1, and k_2 values that are given in Table 32. Water concentration is additionally presented in Figure 25.**



Source: own illustration, Fraunhofer IME

Figure 25: **Detailed plot of the mean water concentrations of fluoxetine during the HYBIT study at a nominal pH of 6. At t = 96hrs the measured concentration was lower than the -20% TWA limit, all other measurements remained stable in a ± 20% TWA range. The measurement at t = 96hrs was not included in the TWA calculation:**



Source: own illustration, Fraunhofer IME

Table 32: BCF parameters for the HYBIT study with fluoxetine (ionised).

Parameter	Value	Unit
k_1	1.806 ± 0.06199	L/(kg*h) L/(kg*d)
k_2	0.08571 ± 0.0104	1/h 1/d
$C_{H,SS}$	0.9876 ± 0.13	mg/kg
TWA	0.04907 ± 0.0065	µg/L
Lipid	3.05 ± 0.42	%
BCF_{SS}	20 ± 3.8	L/kg
BCF_K	21 ± 2.7	L/kg
BCF_{SSL}	20 ± 4.6 (3% lipid) 33 ± 7.7 (5% lipid)	L/kg
BCF_{KL}	21 ± 3.9 (3% lipid) 35 ± 6.5 (5% lipid)	L/kg
Mortality	14.0	%

A.2.7 Fluoxetine (partly neutral)

This chapter describes the results of the fluoxetine HYBIT that was performed at a nominal pH of 9. This pH was facilitated by adding NaOH and TRIS buffer (final concentration 10 mM) to the holding and dilution water of the test facility, which has a natural pH of approx. 7.5 – 8.2.

The HYBIT study with fluoxetine (partly neutral) was conducted successfully, but with restrictions. A clogged tube prevented the stock solution from entering the mixing chamber which caused a drop in the water concentration.

Figure 26 shows the results for tissue and water concentrations and the determined pH values. The visual inspection of the tissue concentrations shows that steady state may have been reached in a time frame ranging from 48 hrs to 144 hrs (end of uptake). Calculations cannot fully confirm that the tissue concentrations lie within a ± 20% range around the mean tissue concentration for that time. The tissue concentrations at t = 144 hrs (end of uptake) spread considerably and range from 9.02 – 27.3 mg/kg with a relative standard deviation of 48.7%. With a mean of 19.03 mg/kg the concentrations at t = 144 hrs are considerably higher than in the remaining time of the assumed steady state (10.0 – 13.9 mg/kg). The tissue concentration in the depuration phase follows the concentration trend set by the tissue concentration at the end of the uptake phase. The high standard deviations persist.

The water concentration is displayed in more detail in Figure 27. The water concentration determined at t = 72 hrs is considerably lower than the lower 20% TWA range. The water concentration measurements prior to t = 72 hrs lie above the upper 20% TWA range. If the value determined at t = 72 hrs is excluded from the TWA calculation, all remaining values lie within the ± 20% TWA range. The determined TWA of 42.47 µg/L is lower than some reported solubility values that all lie in a mg/L range (*cf*. chapter A.2.6 “Fluoxetine (ionised)”).

The mean lipid value determined for *H. azteca* during the fluoxetine (partly neutral) study was 2.82%. The lipid values determined at the start of the uptake phase were considerably lower than the remaining values. Since the data that was determined in the middle and at the end of the study better reflects the lipid content of the amphipods during the study, the lipid content at t= 0hrs was not used to calculate the mean lipid content.

The calculated mortality during the test was 15.3% and thus below the threshold of 20%. Temperature was held steadily around a mean of 23.4°C and differed by max. ± 1°C which is below the allowed ± 2°C. The mean pH value was at 8.9 with minimum and maximum pH values of 8.83 and 8.99. Oxygen saturations were at a mean of 91.8% (≙ 7.51 mg/L) and did not fall lower than 81.1% (≙ 6.68 mg/L). Total flow rates were kept at steady 2.06 L/h throughout the test. Ammonia and nitrite were non- to barely detectable; nitrate values reached a maximum level of 9 mg/L and were thus below any threshold of concern. NPOC levels ranged from 404 – 496 mg/L, which is much higher than the threshold of 2 mg/L. These extraordinary high levels can be explained by the addition of the 10 mM TRIS as buffer substance, as this accounts for 480 mg/L additional carbon in the medium.

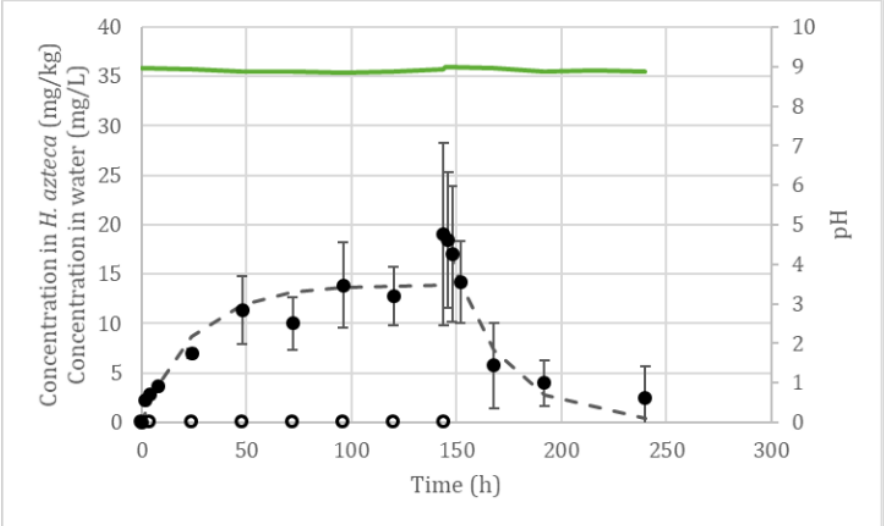
The calculated, non-lipid-normalised BCF values are 280 ± 135.4 (steady state) and 316 ± 158.7 (kinetic). The similarity between steady state and kinetic BCF furthermore indicate that steady state was reached. Table 33 shows the lipid normalised BCF values.

The metabolite norfluoxetine was also measured in the *H. azteca* samples. The concentration of this metabolite increases constantly and does not reach a steady-state. At the end of the uptake phase the concentration of the metabolite is approximately as high as the concentration of the parent substance fluoxetine or lower.

Note:

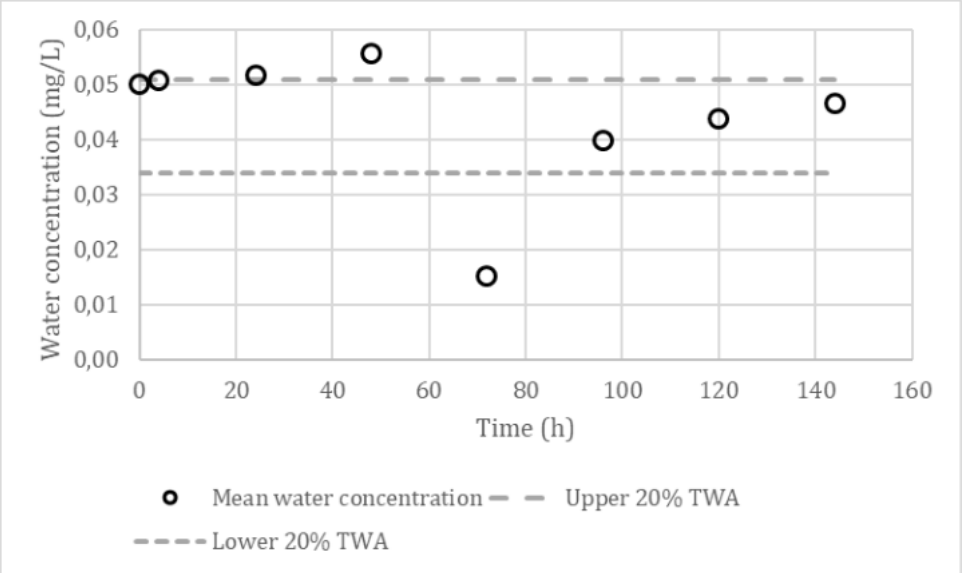
Raw data related to the HYBIT experiments are available as supplementary materials “HYBIT raw data” and “analytics”.

Figure 26: Concentration data of the HYBIT study with fluoxetine (partly neutral). The pH value is presented on the right x-axis. The displayed model fit was calculated from TWA value, k_1, and k_2 values that are given in Table 33. Water concentration is additionally presented in Figure 27.



Source: own illustration, Fraunhofer IME

Figure 27: **Detailed plot of the mean water concentrations of fluoxetine during the HYBIT study at a nominal pH of 9. At t = 72hrs the measured concentration was considerably lower than the -20% TWA limit. Concentrations at the start of the experiment are higher than the + 20% TWA range.**



Source: own illustration, Fraunhofer IME

Table 33: **BCF parameters for the HYBIT study with fluoxetine (partly neutral).**

Parameter	Value	Unit
k_1	13.359 ± 0.907	L/(kg*h) L/(kg*d)
k_2	0.0408 ± 0.0136	1/h 1/d
$C_{H,SS}$	13.41 ± 5.49	mg/kg
TWA	0.0425 ± 0.0124	µg/L
Lipid	2.82 ± 0.65	%
BCF_{SS}	280 ± 135.4	L/kg
BCF_K	316 ± 158.7	L/kg
BCF_{SSL}	336 ± 185.7 (3% lipid) 560 ± 309.5 (5% lipid)	L/kg
BCF_{KL}	348 ± 143.0 (3% lipid) 580 ± 238.4 (5% lipid)	L/kg
Mortality	15.3	%

A.2.8 Overview: validity criteria

Table 34: Overview: validity criteria.

HYBIT	Water temperature	pH	Dissolved oxygen	Stable TWA	Water concentration below solubility	Mortality (< 20%)	[NPOC]
Pyrene	✓	✓	✓	✓	✓	✓	✓
β-HCH	✓	✓	✓	✓	✓	✓	X
UV-234	✓	✓	✓	X	n.d.	✓	✓
Diclofenac (pH 7.8)	✓	✓	✓	✓	n.d.	✓	X
Diclofenac (pH 5)	✓	n.a.	✓	✓	n.d.	X	n.a.*
Fluoxetine (pH 6)	✓	n.a.	✓	X	n.d.	✓	n.a.*
Fluoxetine (pH 9)	✓	n.a.	✓	X	n.d.	✓	n.a.*

n.d. = not determinable

n.a. = not applicable

*values were above calibration (100 mg/L)

A.3 In vitro substrate depletion assays with rainbow trout hepatocytes (OECD TG 319A) – Background information

In silico models have been developed to reduce the amount of animal testing for bioconcentration assessment. Some in silico prediction tools require information on the hepatic biotransformation of a compound. This rate cannot be predicted by the common input parameter for bioaccumulation in silico models, which is the log K_{OW}. Hepatic biotransformation plays an important role in the reduction of the body burden of a substance in an organism. Excluding this information can lead to an overestimation of the BCF. In vitro methods using hepatocytes or hepatic material have long been applied in pharmaceutical science to investigate the fate of pharmaceuticals. These methods have been refined and standardised for hepatocytes and hepatic S9 fraction of rainbow trout and finally adopted as OECD TG 319A/B with an accompanying Guidance Document (OECD 2018a, b, c).

The development of the OECD TG 319A/B was accompanied by a ring trial and several preliminary studies that demonstrated the applicability of the experimental system across a broad set of substances. The data obtained in the ring trial and adjacent studies demonstrated that reproducibility and transferability is given (Nichols et al. 2018a).

The Guidance Document (GD) includes additional information on performing the two OECD Test Guidelines (TG 319A and TG 319B). It suggests a variety of substances that may be used as positive control (OECD 2018c) which involves running assays with a known reference chemical alongside the test chemicals to ensure that the experimental setup is functioning correctly and that the conditions used in the experiment can produce the expected outcome. This allows researchers to verify the enzymatic activity of the biological material by enzyme

characterisation assays (EROD, e.g.) and ensures that the systems are working as intended. Most commonly, the use of pyrene as positive control reference substance is found in the literature. OECD TG 319A/B refers to standardised protocols in the literature for enzyme characterisation assays. While for S9 fractions, these protocols are straightforward, the application in hepatocytes can vary. Most labs use cell lysates to determine enzymatic activities (Fay et al. 2014a, Fay et al. 2017, Nichols et al. 2018a, Nabb et al .2006), but there are also protocols and applications using intact cells (Laville et al. 2004, Smeets et al. 2002),

Testing of neutral organic chemicals

The ring trial for OECD TG 319A/B was performed with a set of 6 different substances with a log K_{OW} range of approx. 4.09 – 6.2 (Nichols et al. 2018a). Other studies included a set of fragrances (Laue et al. 2014), a substance class that is characterised by volatile characteristics. The tested fragrances had a log K_{OW} range of approx. 4.0 – 5.7. The dataset was further enhanced by data for additional chemicals (e.g., Laue et al. 2020, Kropf et al. 2020, Laue et al. 2023).

Lower hydrophobic compounds were tested, e.g., molinate with a consensus log K_{OW} of 2.7 (Han et al. 2007, Han et al. 2009, Mingoia et al. 2010) and atrazine (Han et al. 2007) with a consensus log K_{OW} of 2.5. Also, metabolisable substances in the lower to mid log K_{OW} range (Kosfeld et al. 2019), and additional PAH data that also explored the effects of starting concentration, solvent concentration and incubation time on the determined reaction rate (Nichols et al. 2018b) were tested. Comparisons between S9 and hepatocytes as test systems were also conducted (Fay et al. 2014b, Fay et al. 2017, Han et al 2009).

The Guidance Document for the OECD TG 319 A/B includes an IVIVE model developed by Nichols et al. 2013a. Models with similar assumptions and variations of this model have been described in Han et al. (2007), Han et al. (2009) and Krause & Goss (2020) and an updated version of the Nichols et al. 2013 model has been implemented in the EAS-E suite. Laue et al. (2023) provided a regression-based approach to the BCF-IVIVE. Overall, the IVIVE-BCFs show good correlation with the respective in vivo fish BCFs, but a look at the literature data shows that the use of different IVIVE models or different parameter settings within an IVIVE model can lead to a high variability of the results obtained. In those cases, expert judgement and, if possible, a recalculation of the IVIVE BCFs using more modern models may be necessary to assess the validity and plausibility of the tests and the resulting bioaccumulation potential.

In the literature, the applicability of the substrate depletion assay protocols was also evaluated for several substance classes, as described in the following chapters.

Testing hydrophobic and volatile substances

Substances with elevated hydrophobicity and/ or volatility can pose challenges for exposure. They can either bind to the surface of the incubation vessel or disappear from the incubation vessel by vaporisation. Still, many substances with volatile characteristics can be tested by using the standard dosing approach (e.g. Laue et al. 2014, Laue et al. 2020, Kropf et al. 2020, Cantu et al. 2025). Still, thin film or sorbent dosing techniques were evaluated in some studies and compared to solvent spiking methods for ensuring reliable exposures in the in vitro depletion assays, if required (Lee et al. 2014, Lo et al. 2015). Passive dosing can pose additional modelling effort, as not only the depletion rate, which is the result of the biotransformation by the hepatic material has to be determined, but also the mass-transfer rate constants that describe the interactions between the test substance and the sorbent-phase (e.g., Smith et al. 2012). Passive dosing was also used to prepare the dosing solutions for a depletion assay, providing another option for the preparation of test solutions (Kropf et al. 2020).

Testing of ionisable organic compounds

In contrast to in vivo BCF testing, it is not sensible to perform in vitro testing at different pH levels. The hepatic enzymes are not only temperature, but also pH-optimised and thus, the pH demanded for incubation in the TG should not be altered. Otherwise, it cannot be ensured that the hepatic enzymes can work in an optimised state and the determined biotransformation is likely underestimated.

The testing of ionisable substances was performed in large quantities. Sets of ionisable compounds, mainly pharmaceuticals, were tested in vitro. For some compounds, the impact of chirality on biotransformation was also investigated (Connors et al. 2013).

While the in vitro depletion assay may be straightforward, modelling their distribution in silico requires additional effort. Armitage et al. 2017 summarises the aspects that need consideration when modelling the bioaccumulation of IOCs in fish. Extrapolation models with a focus on ionisable compounds were developed (Armitage et al. 2013) and can be used within the EAS-E Suite but have not fully been shared as stand-alone application unlike the Nichols et al. 2013a or the Krause and Goss 2020 models, which are available as Excel spreadsheets for the extrapolation of both, S9 and hepatocyte depletion data. Additional input parameters such as for example a storage lipid-water partition coefficient (log K_{SW}), or the p*K*a value of the substance are necessary for the IVIVE of ionisable substances (*cf*. "Introduction to the BIONIC model V1 and V2") and should ideally be deduced with similar considerations as the log K_{OW}.

Testing of surfactants

Surfactants pose another challenge for bioaccumulation testing due to their unique partitioning characteristics, as chapter 4.1.3 of this report has pointed out. Cationic and anionic surfactants were subjected to in vitro depletion assay testing using rainbow trout S9 (Chen et al. 2016). Droge et al. 2021 mention that there may be yet undefined rate limitations for the diffusion of cationic surfactants across gill and hepatocyte membranes. These aspects need to be considered for future IVIVE applications. Model development for in silico QSARs can directly benefit from in vitro depletion assays. Biotransformation of some of the tested surfactants could be determined.

Testing of PFAS

To our knowledge, PFAS have thus far not been tested using the OECD TG 319A/B. Instead of measuring the biotransformation to obtain a loss rate for the parent substance, Butt et al. (2010) investigated the biotransformation of a precursor compound (PFCA) 8:2 FTAc to 8:2 FTOH in rainbow trout-S9. This compound can be further transformed to perfluorooctanoate and perfluorononanoate. The S9 test system, however, probably lacked the required co-factors for this transformation step. PFOS is considered a final biotransformation product that can be generated from a large set of precursor substances, regardless of species (Kolanczyk et al. 2023). In future applications, in vitro assays could hence provide insight into processes such as the catabolism of substances from precursor substances that can also contribute to bioaccumulation.

Additional options during testing

OECD TG 319A/B provides the description of a substrate depletion assay using rainbow trout hepatocyte of hepatic S9 fractions of rainbow trout. Literature shows that in addition to substrate depletion rates, metabolite profiles of the test substance can also be determined. Examples are the fungicides azoxystrobin, prochloraz and trifloxystrobin, the herbicide terbutryn (Kosfeld et al. 2020), as well as the pharmaceutical diclofenac (Fu et al. 2020). In vitro metabolite profiles were similar to in vivo metabolite profiles for the pesticide methoxychlor (Bischof et al. 2016). Alterations to the OECD 319A/B test procedure can involve larger

incubation volumes, higher starting concentrations and longer incubations to maximise the metabolite diversity at a quantifiable level (Kosfeld et al. 2020, Bischof et al. 2016). Analytical procedures need to be developed to detect and quantify the metabolites. The experiment can also involve ^{14}C-radiolabel testing.

A.4 In vitro substrate depletion assays with rainbow trout hepatocytes (OECD TG 319A) – Material and Methods

A.4.1 Cellular material

Rainbow trout hepatocytes were obtained from in-house sources and were prepared from rainbow trout reared in the hatchery of Fraunhofer IME-AE. The cells were isolated and stored using established and published protocols (*cf*. Bischof et al. 2016; Kosfeld et al. 2020). Basic metabolic activities were determined using the EROD, GSH and UGT assay, respectively (Nabb et al. 2006, Habig et al. 1974, Castren and Oikari, 1983; Ladd et al. 2016).

In total, three different cell lots were used for different tests. All details regarding the donor fish and the results of the characterisation are summarised in Table 35 and Table 36, respectively.

Table 35: Details for the rainbow trout hepatocyte lots used in this project.

	Lot HEP-14-2023	Lot HEP-11-2021	Lot VK 1
Fish	Rainbow trout (*Oncorhynchus mykiss*)	Rainbow trout (*Oncorhynchus mykiss*)	Rainbow trout (*Oncorhynchus mykiss*)
Source fish	Fischzucht Störk, Bad Saulgau	Fischzucht Störk, Bad Saulgau	Fischzucht Störk, Bad Saulgau
Source HEPs	Fraunhofer IME in-house preparation	Fraunhofer IME in-house preparation	Fraunhofer IME in-house preparation
Pool?	Yes (5 fish)	Yes (5 fish)	Yes (4 fish)
Gender	Mixed	Female	Mixed
Weight	232 g (mean)	416 g (mean)	285 g (mean)
Length	26.7 cm (mean)	32.2 cm (mean)	29.1 cm (mean
Age	Approx. 1 year	Approx. 2 years	Approx. 1.5 years

Table 36: Results of the enzymatic characterisation assays for the rainbow trout hepatocyte lots used in this project.

Assay	Lot HEP-14-2023	Lot HEP-11-2021	Lot VK 1
EROD (pmol/ min/ mg protein)	0.32	0.53	0.42
GSH (nmol/ min/ mg protein)	510.0	432.8	863.5
UGT (nmol/ min/ mg protein)	440.8	375.1	563.1

A.4.2 Test setups

Preliminary experiments were performed to identify suitable incubation conditions. The final settings for the main depletion experiments are summarised in Table 37. Samples were extracted via vortexing for 10 minutes and samples were frozen at ≤ -18°C over night. Before a suitable aliquot (200 – 300 µL) of the extract was transferred into a GC-vial with fused insert for analysis, a centrifugation at 20000 x g at 4°C was done. For pyrene and β-HCH samples were centrifuged at room temperatures and less g (exact values not available).

Table 37: Test conditions for the in vitro depletion assays using rainbow trout hepatocytes.

	Pyrene	Pyrene	β-HCH	β-HCH	UV-234
Cellular material	RT-HEPs	RT-HEPs	RT-HEPs	RT-HEPs	RT-HEPs
Lot name	HEP-14-2023	HEP-14-2023	HEP-11-2021	VK1	HEP-14-2023
Reaction buffer	L-15 medium at pH 7.8	L-15 medium at pH 7.8	L-15 medium at pH 7.8	L-15 medium at pH 7.8	L-15 medium at pH 7.8
Cellular concentration	2 * 10^6 cells/mL	2 * 10^6 cells/mL	2 * 10^6 cells/mL	2 * 10^6 cells/mL	2 * 10^6 cells/mL
Test approach	Multiple vial	Multiple vial	Multiple vial	Multiple vial	Single vial
Runs/ replicates	1	2	2	2	3
Sampling times	0, 5, 10, 20, 30, 50 min	0, 2, 5, 10, 20, 30 min	0, 15, 30, 60, 120, 180, 240 min	0, 15, 30, 60, 120, 180, 240 min	0, 15, 30, 60, 120, 180, 240 min
Reaction Volume	500µL	500µL	500µL	500µL	1000µL
Spike Volume	2.5µL	2.5µL	2.5µL	2.5µL	5µL
Aliquot Volume	500µL	500µL	500µL	500µL	100µL
Stopping Volume	300µL	300µL	300µL	300µL	400µL

	Pyrene	Pyrene	β-HCH	β-HCH	UV-234
Stopping Solution	Acetonitrile + IS and 200µl n-hexane	Acetonitrile + IS and 200µl n-hexane	Methanol + IS and 200µl n-hexane	Methanol + IS and 200µl n-hexane	Acetonitrile + IS
Internal standard	Pyrene-d10	Pyrene-d10	HCB	HCB	UV-234-d4
Nominal start concentration	0.04µM/L	0.04µM/L	0.05 µM/L	0.05 µM/L	0.03µM 0
Spiking solution	Acetone	Acetone	Acetone	Acetone	Acetonitrile
Reaction temperature	12°C	12°C	12°C	12°C	12°C
Reaction vessel	1.8mL GC-Vial (clear glass) with PTFE septum	1.8mL GC-Vial (clear glass) with PTFE septum	1.8mL GC-Vial (clear glass) with PTFE septum	1.8mL GC-Vial (clear glass) with PTFE septum	7-mL scintillation vial, loosely capped
Analytical method	GC-MS	GC-MS	GC-MS	GC-MS	LC-MS

The supervision of the analytical work and the calculation of substance concentrations in media samples collected during the In vitro experiments were carried out by Dr. Bettina Dudek and Dr. Carolin Mörtl (UV-234), and Claire MacKenzie (Pyrene, ß-HCH), all Fraunhofer IME, Schmallenberg.

A.4.3 Statistics

Slopes of the linear regressions were compared using GraphPad Prism (Version 6.07 for Windows). Data was entered into datasheets. Afterwards, the analysis function “Fit a line with linear regression” was chosen, activating the sub option “Test whether slopes and intercepts are significantly different”. The slopes of active runs were compared against each other individually and all together. Slopes of active and HI incubations were checked whether they are significantly different from zero. A statistical difference is given if the calculated p is < 0.05.

If active runs were detected that are not significantly different from zero and/ or their HI run, it was concluded that there is no detectable biotic depletion. Accordingly, the $Cl_{int,\ in\ vitro}$ was set to 0.

A.4.4 In vitro to in vivo extrapolation (IVIVE)

It is possible to use the in vitro depletion rates that are determined in the experiments to refine in silico bioconcentration predictions. The IVIVE model by Krause and Goss with the name “B-compass Fish” (Krause & Goss 2020) was used in this project to extrapolate the determined clearance rates to bioconcentration factors. Version 1.2 of the spreadsheet was downloaded from the UFZ homepage and was used without alterations of the settings (Krause, S. and K.-U. Goss, B-compass Fish Version 1.2 (K_{OW} approach), Leipzig, Germany, Helmholtz Centre for Environmental Research, 2021). Input parameters were chosen as follows:

- Log K_{OW}: Consolidated log K_{OW} of the test substance (Nendza et al. 2025)
- Option a) "hepatic biotransformation": "cell assay"
- K_{assay}: Mean $CL_{int, in\ vitro}$ (mL/h/10^6 cells) determined in the assay of this project
- $C_{viable\ cells\ in\ assay}$: 1 *$10^6$ cells/mL_{assay}
- Viability: Mean viability determined in the assay of this project
- V_{assay}: 1 mL_{assay}
- Option b) "gill biotransformation": "N/A"
- Option c) "GIT biotransformation": "N/A"
- Bodyweight of modeled fish: 10 g (default setting)
- Temperature: 12°C (adjusted to assay temperature)
- Lipid content of modeled fish: 0.05 (default setting)

Since default settings were used, the model by Sijm et al. 1995 was used to calculate the gill uptake rate constant k_1.

Depletion rates reported in the literature have also been extrapolated to BCFs using the B-compass Fish model. If these depletion rates were obtained from S9 fractions, all input parameters were chosen as described above with the following exceptions: Setting a) was switched to "S9 assay" and the input parameters for the S9 assay were the following:

- K_{assay}: $CL_{int, in\ vitro}$ (mL/h/mg protein) determined in the literature
- $C_{S9\ in\ assay}$: 1 mg_{S9} / mL_{assay}
- V_{assay}: 1mL

Other IVIVE models are also available, most prominent is the IVIVE model by Nichols et al. (2013a) that is suggested in the Guidance Document of OECD TG 319A/B. This model has been refined in the past years, as described in Laue et al. (2023). The same publication also introduces a BCF extrapolation model that relies on a regression approach. A comparison of the BCF extrapolations obtained with the regression model against the refined Nichols model and the B-compass fish model is provided. The dataset used in the comparison shows that all three models deliver similar outputs and that the selection of the method to calculate the uptake rate k_1 impacts the output of the two IVIVE models. The selection of the method to calculate k_1 in the IVIVE models appears to impact substances with elevated log K_{OW} values and low biotransformation rates. Since the revised Nichols model is not yet available as spreadsheet and the Laue et al. (2023) regression model is focussed on S9 only, the B-compass Fish model was selected for all IVIVEs reported this project. It provides input options for both, S9 and hepatocyte depletion rates and can thus provide a unified extrapolation basis for all depletion rates evaluated in this project.

In this project, some ionisable substances were tested. The B-compass Fish model is not applicable to ionisable compounds. An IVIVE model for ionisable compounds, called BIONIC (*cf.* Armitage et al. 2013) has been developed as part of the Cefic LRI ECO37 project. Among log K_{OW}, other input parameters are required. It is out of scope of this project to evaluate the derivation

of the required partitioning information and to also evaluate their influence and the influence of the model settings on the IVIVE BCFs.

As a shortcut, the B-compass Fish model was used, despite it being out of applicability domain. The log D values of the substances at a neutral and at an ionised state were used as input for the model. While this procedure ignores physiological processes that are specifically relevant for ionisable compounds, it may provide worst- and/ or best-case scenarios for the bioconcentration of the compounds. The resulting IVIVE BCFs must be interpreted with caution.

A.4.5 Preliminary experiments

β-HCH

The analytical method for the analysis of β-HCH in rainbow trout hepatocyte extracts was based on Lee at al. (2022) with minor modifications. The extraction had to be fine-tuned. 200 µL of n-hexane were used to extract the cellular material, but n-hexane alone was not able to inactivate the active cells reliably. Therefore, it was decided to inactivate the cells by other means:

1. Using ice-cold acetonitrile as stopping solution (stored at ≤ -18°C)
2. Adding KCl for an improved phase separation
3. Inactivating cells by snap-freezing
4. Inactivating cells using heat-inactivation
5. Using ice-cold methanol as stopping solution

In optimization trials 1) – 4) a gelatinous mid-phase formed in the hepatocyte extract, optimization trial 1) produced a three-phased extract. The sensitivity of the analytical method was considerably reduced, presumably because some β-HCH was stuck in the ACN phase and did not move into the n-hexane phase. Extraction method 5) was performed with shorter vortexing times and was provided acceptable analytical results. All fine tuning was performed using heat inactivated hepatocytes and were prepared in the same manner as matrix calibration samples.

A.4.6 UV-234

Since UV-234 is a highly hydrophobic substance that had proven itself to be challenging to disperse in water, some preliminary experiments were performed to understand the substance's behavior in the test system.

In a first experiment it was evaluated whether it is possible to obtain samples suitable for a matrix calibration using L-15 medium only. In the same experiment, it was evaluated whether the time that the test substance is kept within the incubation vessel has an influence on the quality of the calibration curve. The results indicated that a calibration with matrix samples could be possible. Incubation time was no apparent factor for the recoveries of the test substance.

This experiment was repeated with heat-inactivated (HI) rainbow trout hepatocyte material. The recoveries improved slightly and again, there was no apparent time-dependent influence on the recovery. Next, it was evaluated whether active rainbow trout hepatocytes are able to metabolise UV-234. Three incubations with active hepatocytes were performed, each with a different starting concentration (125 µg/L, 37.5 µg/L, 12.5 µg/L). Additionally, a negative

control at the middle starting concentration was performed. Samples were drawn after 30, 60, 120, 180 and 240 minutes. No substance loss was visible in the negative control and at the mid-concentration with active hepatocytes. A minor loss of substance was seen in the incubation with the high and the low starting concentration.

To evaluate whether this substance loss was due to biotic or abiotic processes, the last experiment was repeated, but this time the incubation vessels had been pre-equilibrated over night with L-15 medium that had been spiked with the test substance. The incubation was initiated by removing the L-15 medium and adding the active or HI hepatocytes, that were then spiked with the test substance. The results are scattered, but similar trends as in the previous experiment were seen.

A follow-up experiment was performed to check whether desorption effects occur when pre-equilibrated incubation vessels are used. In theory it could be possible that desorbed substances masks substance loss. The test was conducted with the mid-concentration range. Incubation vessels were pre-equilibrated using L-15 medium that was spiked with the test substance. The spiked medium was removed from the test vessels and non-spiked medium and non-spiked HI-hepatocytes were added to the vessels, respectively. After 30 minutes, UV-234 can be detected in the L-15 medium and in the HI-hepatocyte solution. Again, scatter is visible and thus the results are indicative only. While the UV-234 content in the samples with L-15 medium shows an apparent reduction over time, the detectable UV-234 concentration in the test vessels with HI-hepatocytes shows a slight apparent increase. However, no more than 6 – 14% of the nominal starting concentration was detected in the vessels and thus were caused by desorption.

This test was repeated with three different starting concentrations. In addition to the desorption control vessels that contained either L-15 medium or HI-hepatocytes, a third option was added: HI-hepatocytes were spiked with the test substance at the same concentration as the nominal concentration for equilibration. Each treatment was set up in duplicates. The results indicated that vessels that had been equilibrated at higher concentrations release larger amounts of substance into non-spiked medium. There are individual samples with unexpectedly high concentrations and the duplicates of each treatment do not always match. Some of the “re-spiked” treatments remain at levels approximately stable around 100% of the nominal concentrations, others drop in their concentrations at first but slowly recover over the course of the 4 hrs of incubation.

In a final preliminary experiment, four different treatments were compared: L-15 medium, HI-S9 fraction of rainbow trout, HI-hepatocytes and active hepatocytes were spiked with an acetonic UV-234 spiking solution and all four treatments were incubated for 4 hrs with samplings at 0, 30, 60, 120, 180 and 240 minutes. The vessels had not been pre-equilibrated beforehand. All treatments were set up in duplicates. The results show that HI-S9, HI-hepatocytes and live hepatocytes display very similar concentrations and remain at a stable concentration level. The medium only treatment showed lower concentrations with comparatively higher fluctuations.

The conclusions from the preliminary experiments were the following: UV-234 apparently adheres to the test vessel surface, but probably also to the hepatic matrix. The test substance probably also desorbs from the test vessel’s surface. Especially the latter process was shown to be dependent on the concentration applied in the test vessel. The sorption and desorption processes, however, appear to be either neglectable or are likely not statistically detectable with the methods applied. Incubations with active hepatocytes in the preliminary experiments showed that there was no detectable, systematic, constant loss of substance. It was decided to perform the main tests in the usual manner by spiking non-equilibrated, freshly prepared hepatocyte solutions.

A.5 In vitro depletion assays with rat hepatocytes

A.5.1 Literature background

An efficiently absorbed, non-biotransformed neutral organic substance with a log $K_{OA} \geq 5$ in combination with a log $K_{OW} \geq 2$ is considered to potentially biomagnify in vertebrates of terrestrial food chains and air-breathing marine wildlife as well as in humans (ECHA 2023a; ECHA 2022), while the substances with log $K_{OW} < 2$ have a reduced gastrointestinal uptake or are efficiently excreted in urine, and therefore do not biomagnify even though their K_{OA} is high (Armitage and Gobas, 2007; Kelly et al., 2007; Gobas et al., 2009; McLachlan et al., 2011; Goss et al., 2013). However, among the multitude of organic chemicals, only a very small fraction is sufficiently volatile for respiration to be an efficient elimination pathway. Screening methods to determine the bioaccumulation potential of chemicals in air-breathing species need to include an assessment of the rate of chemical transformation or metabolism (Gobas et al. 2003). Wania et al. 2022 showed the potential of a screening method including predicted biotransformation half-lives to identify organic chemicals which are not subject to bioaccumulation in air-breathing organisms. Recently, an in vitro screening approach for assessing bioaccumulation of neutral hydrophobic organic chemicals from mammalian liver S9 biotransformation assays was developed and tested (Lee et al. 2022). It was shown that a substance is not expected to biomagnify in rats ($BMF_L<1$) when the in vitro biotransformation rate constant measured in the rat liver S9 assay is greater than approximately 0.3 h^{-1}. In vitro-derived total elimination rate constants were in good agreement with reported in vivo data for selected test chemicals with available in vivo data (Lee et al. 2022). Previously, methods for in-vitro hepatic clearance studies were developed. Black et al. (2021) evaluated and compared in vitro intrinsic clearance rates measured using cryopreserved hepatocytes from humans, rats, and rainbow trout and described protocols for the conduction of depletion assays. Louisse et al. (2020) compared the experimental conditions of different depletion assays that are reported in the literature and their respective in vitro hepatic clearance data to identify the key experimental conditions that influence the clearance rate which is essential with respect to harmonisation of test methods for in vitro hepatic clearance studies.

A.5.2 Preceding considerations

While for trout it is stated that the quality of depletion rates is not expected to be influenced by the choice of test system, i.e., S9 or hepatocytes (Fay et al. 2017). A comparison of rat S9 fractions against rat hepatocytes regarding the test substance stability was conducted with a dataset of 456 substances (Richardson et al. 2016). This comparison that is part of a method optimisation in pharmacology, shows that the majority of substances (approx. 70 - 80%) show a comparable stability in rat and human S9 fractions and hepatocytes. Depending on the substance, the stability may be higher in hepatocytes due to poor permeability of the test substance into the cells. The guidance document for OECD TGs 319A/B states that trout hepatocytes may be preferred over S9 fraction in cases where it is known that minor phase II enzymes are directly involved in the biotransformation of the parent substance, as the required co-factors are not commonly added to the S9 incubation mix (OECD 2018c). The Guidance Document also states that hepatocytes are able to maintain their biotransformation integrity for a longer period compared to S9 fractions (OECD 2018c). The Guidance Document also states that the incubation time for S9 fractions may be prolonged to up to 4 hrs for slowly biotransformed

test items. The determination of the exact lifetimes of both in vitro test systems may benefit from more research.

For this project we assume that similar effects are present for hepatic rat material, hence it was decided to perform the rat depletion assays for this project using rat hepatocytes. The publication of Black et al. (2021) was used as inspiration for the assay conditions with the following aspects noted for the Fraunhofer-IME protocols:

- Incubation medium: William's E Medium
- Cell density: 0.5 – 1 x 10^6 cells/mL
- Pre-incubation for 10 minutes at 37°C
- Incubation medium pH: 7.4

The review of Louisse et al. (2020) confirmed these settings as suitable for rat hepatocytes.

A.5.3 Cellular material

Rat hepatocytes were retrieved from Primacyt and were shipped alongside the required media for thawing and incubation. EROD enzyme characteristics were performed by Primcyt and resulted in 86.2 ± 2.6 pmol/ (mg * min). Post-thawing viability should reach at least 74.9 % according to the information for lot "RH191205 Pool".

Table 38: Details of the rat hepatocyte lot "RH191205 Pool"; Excerpt of the CoA by the supplier Primacyt, Schwerin.

Criterion	Details/ Results	
Species	Rattus norvegicus forma domestica	
Strain	CD IGS rat (Sprague Dawley) // Crl:CD (SD)	
Pool?	Yes, 5 animals	
Gender	Male	
Post-thaw viability	74.9 ± 5.6 % (n = 2)	
Post-thaw yield per vial	6.2 ± 0.0 x 106 cells (n = 2)	
Recovery	62% (n = 2)	
CYP P450 activity in culture after thawing:	pmol/(mg × min)	X-fold induction
Ethoxyresorufin-O-deethylation:	24well: 86.2 ± 2.6 (n=2)	21.3
Induction with 25 µM ß-Naphthoflavone	96well: 326.0 ± 48.9 (n=1)	22.1

A.5.4 Test setups

Preliminary experiments were performed to identify suitable incubation conditions. The final settings for the main depletion experiments are summarised in Table 39. Samples were extracted via vortexing for 10 minutes and samples were frozen at ≤ -18°C over night. Before a suitable aliquot (200 – 300 µL) of the extract was transferred into a GC-vial with fused insert for analysis, a centrifugation at 20000 x g at 4°C was done.

The supervision of the analytical work and the calculation of substance concentrations in media concentrations collected during the in vitro experiments were carried out by Dr. Bettina Dudek and Dr. Carolin Mörtl (UV-234, Chlorpyrifos) and Claire MacKenzie (Pyrene, ß-HCH), all Fraunhofer IME, Schmallenberg.

Table 39: Test conditions for the in vitro depletion assays using rat hepatocytes.

	Pyrene	Chlorpyrifos	UV-234	UV-234
Cellular material	Rat hepatocytes	Rat hepatocytes	Rat hepatocytes	Rat hepatocytes
Lot name	RH191205 Pool (Primacyt)	RH191205 Pool (Primacyt)	RH191205 Pool (Primacyt)	RH191205 Pool (Primacyt)
Incubation medium	William's E Medium at pH 7.4	William's E Medium at pH 7.4	William's E Medium at pH 7.4	William's E Medium at pH 7.4
Cellular concentration	1 * 10^6 cells/mL	1 * 10^6 cells/mL	1 * 10^6 cells/mL	2 * 10^6 cells/mL
Test approach	Multiple vial	Single vial	Single vial	Single vial
Runs/ replicates	3	2	2	2
Sampling times	0, 5, 10, 20, 30, 45 minutes	0, 5, 15, 30, 60, 90, 120 minutes	0, 15, 30, 60, 90, 120, 180, 240 minutes	0, 5, 10, 20, 30, 45, 60 minutes
Reaction Volume	500 µL	1000µL	1000µL	1000µL
Spike Volume	2.5µL	5µL	5µL	5µL
Aliquot Volume	500µL	100µL	100µL	100µL
Stopping Volume	300µL + 200µL	400µL	400µL	400µL
Stopping Solution	Acetonitrile + n-hexane	Acetonitrile	Acetonitrile	Acetonitrile
Internal standard	Pyren-d10	Chlorpyrifos-(dietyhl-d10)	UV-234-d4	UV-234-d4
Nominal start concentration	0.04µM	0.04µM	0.03µM	0.001µM
Spiking solution	Acetone	Acetonitrile	Acetonitrile	Acetonitrile
Reaction temperature	37°C	37°C	37°C	37°C

	Pyrene	Chlorpyrifos	UV-234	UV-234
Reaction vessel	1.8mL GC-Vial (clear glass) with PTFE septum	7-mL scintillation vial, loosely capped	7-mL scintillation vial, loosely capped	7-mL scintillation vial, loosely capped
Analytical method	GC-MS	LC-MS	LC-MS	LC-MS
Comment		Both runs were performed in parallel using the same HI-control		Both runs were performed in parallel using the same HI-control

A.5.5 Statistics

Slopes of the linear regressions were performed using GraphPad Prism (Version 6.07 for Windows). Data was entered into datasheets. Afterwards, the analysis function "Fit a line with linear regression" was chosen, activating the sub option "Test whether slopes and intercepts are significantly different". The slopes of active runs were compared against each other individually and all together. Slopes of active and HI incubations were checked whether they are significantly different from zero. A difference is present if the calculated p is < 0.05.

If active runs were detected that are not significantly different from zero and/ or their HI run, it was concluded that there is no detectable biotic depletion. Accordingly, the $Cl_{int,\ in\ vitro}$ was set to 0.

A.5.6 In vitro to in vivo extrapolation (IVIVE)

It is possible to use the in vitro depletion rates that are determined in the experiments to refine in silico biomagnification predictions, as shown in Lee et al. (Lee et al. 2022). The IVIVE model used in that publication in applicable to S9 fractions, only. The IVIVE model by Krause and Goss (Krause & Goss 2021) with the name "B-compass Rat" was used in this project since it also enables an IVIVE based on depletion rates determined using hepatocytes.

Version 1.0 of the spreadsheet was downloaded from the UFZ homepage and was used without alterations of the settings (Krause & Goss 2021). Input parameters were chosen as follows:

- Chemical concentration in food: 0.1 $g_{chemical}/g_{food}$ (default value)
- Log K_{OW}: Consolidated log K_{OW} of the test substance (Nendza et al. 2025)
- Log K_{OA}: Consolidated log K_{OA} of the test substance (Annex A.11)
- Option a) "hepatic biotransformation": "cell assay" selected
- K_{assay}: Mean $Cl_{int,\ in\ vitro}$ (mL/h/10^6 cells) determined in the assay of this project
- $C_{viable\ cells\ in\ assay}$: 1 *$10^6$ cells/mL_{assay}
- Viability: Mean viability determined in the assay of this project
- V_{assay}: 1 mL_{assay}
- Option b) "lung biotransformation": "N/A"

- Option c) “GIT biotransformation”: “N/A”
- Option d) “kidney biotransformation: “N/A”
- Bodyweight of modelled rat: 200 g (default setting)

Contrary to the Lee et al. (2022) calculations, the input parameters log K_{OA} and log K_{OW} were not corrected for temperature.

Depletion rates reported in the literature have also been extrapolated to BMFs using the B-compass Rat model (Krause & Goss 2021). If these depletion rates were obtained from S9 fractions, all input parameters were chosen as described above with the following exceptions: Setting a) was switched to “S9 assay” and the input parameters for the S9 assay were the following:

- K_{assay}: $CL_{int,\ in\ vitro}$ (mL/h/mg protein) determined in the literature
- $C_{S9\ in\ assay}$: 1 mg_{S9} / mL_{assay}
- V_{assay}: 1mL

The Lee et al. (2022) IVIVE model was designed for S9 data, accordingly, the hepatocyte depletion rates cannot be extrapolated with that model. The B-compass Rat allows the insertion of S9 or hepatocyte depletion rates and thus provides a unified evaluation basis for the depletion rates evaluated in this project.

In this project, ionisable substances were tested. The B-compass Rat model is not applicable to ionisable compounds. Despite it being out of applicability domain, the log D values of the substances at a neutral and at an ionised state were used as input for the model. While this procedure ignores physiological processes that are specifically relevant for ionisable compounds, it may provide worst- and/ or best-case scenarios for the biomagnification of the compounds. The resulting IVIVE BMFs must be interpreted with caution.

A.6 Results – In vitro substrate depletion assays with rainbow trout hepatocytes (OECD TG 319A)

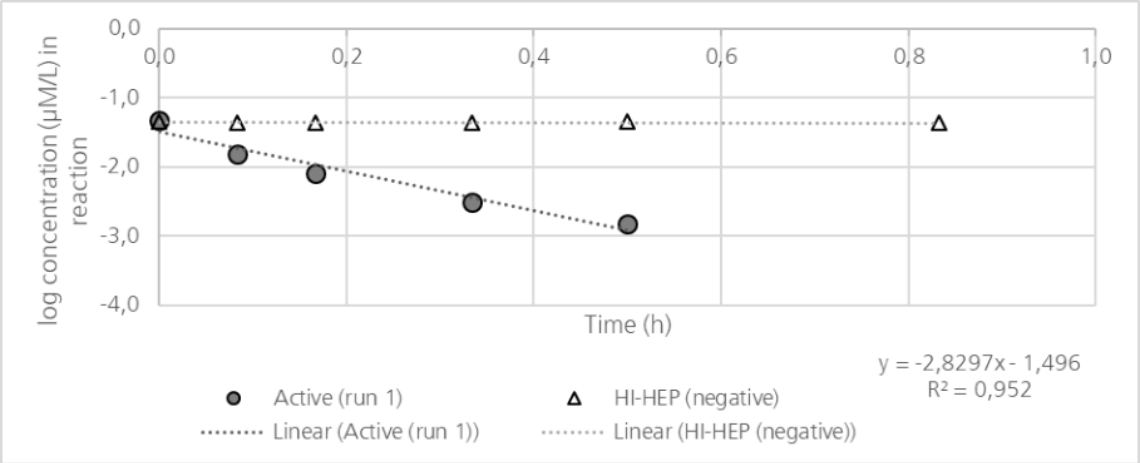
A.6.1 Pyrene

Three runs to test the depletion of pyrene were performed with rainbow trout hepatocytes (Table 40). The first run lasted 50 minutes, based on the incubation duration applied in the ring trial for pyrene depletion assays with rainbow trout hepatocytes. At t = 50 minutes the measured concentration was below LOD and did not differ much from the measurement at t = 30 minutes, which also was below LOD. For the remaining two runs the incubation time was reduced to 30 minutes. Still, the last measured concentrations were below the LOD. Including the measured concentration at t=30 minutes did not impact the R^2 of the linear regression, so it was included despite being below the LOD (Figure 28-Figure 30).

Table 40: Summary of the results of the in vitro depletion assays for pyrene with rainbow trout hepatocytes.

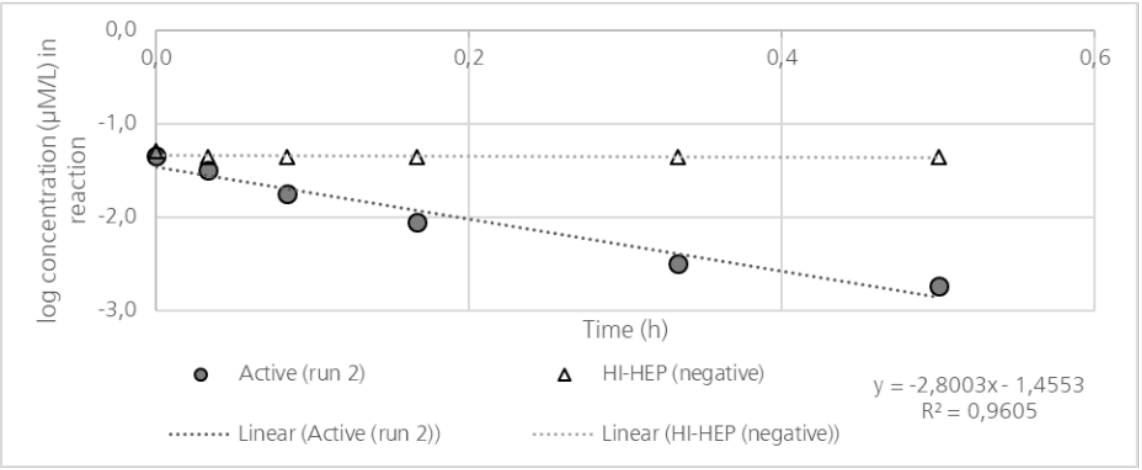
Parameter	Results	Results (calculations only including > LOD)
Runs	3	3
Cell lot	HEP-14-2023	HEP-14-2023
Slope active runs ≠ 0?	Yes (p = < 0.0001 – 0.0045)	Runs 2 & 3: Yes (p = 0.0405 – 0.0448) Run 1: No (0.1613)
Runs comparable?	Yes (p = 0.6344)	Yes (p = 0.9697)
HI slopes ≠ 0?	No (p = 0.1169 – 0.5199)	No (p = 0.1204 – 0.5485)
HI loss of substance < 20%	Yes	Yes
Cellular concentration (mean ± SD)	1.9 ± 0.20 * 10^6 cells/ mL	1.9 ± 0.20 * 106 cells/ mL
Viability (mean ± SD)	83.8 ± 1.5%	83.8 ± 1.5%
Recovery (mean ± SD)	31. 7 ± 3.8 %	31. 7 ± 3.8 %
LOD	0.5µg/L = 0.00247 µM/L	0.5µg/L = 0.00247 µM/L
$CL_{in\ vitro,\ int}$ (mean ± SD)	3.59 ± 0.55 mL/h/10^6 cells	4.16 ± 0.43 mL/h/10^6 cells
Comment	For the first run only 5 datapoints are available. The datapoint t = 50minutes was omitted from the dataset. In all 3 runs the last measured concentration at t=30 minutes was < LOD, but was included in the regression, since it did not impact the R^2.	For the first run only 4 datapoints are above the LOD. The linear regression results in a slope that is not statistically different from zero. For runs 2&3 only 5 datapoints are above the LOD. The slopes of all 3 active incubations are not significantly different from each other.

Figure 28: In vitro depletion assay for pyrene with rainbow trout hepatocytes - Run 1.



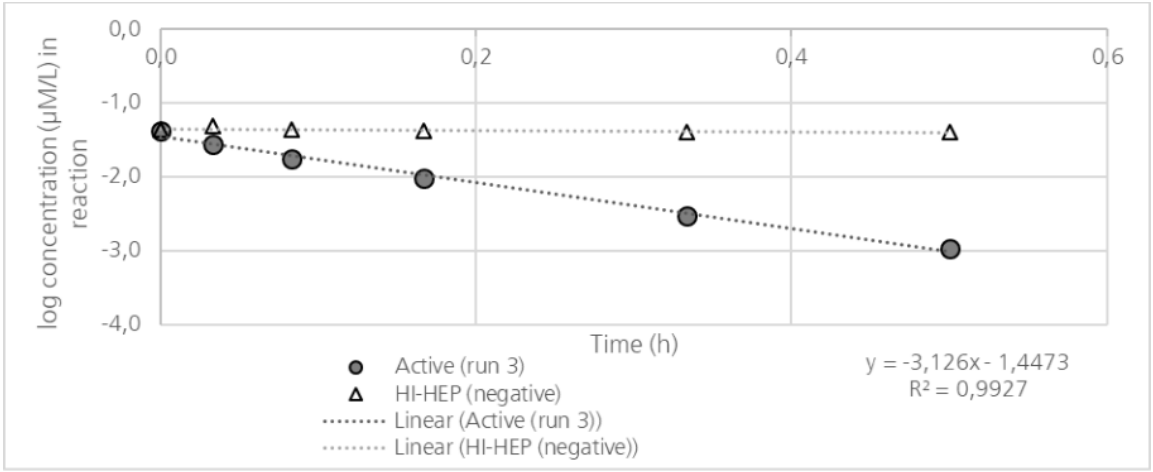
Source: own illustration, Fraunhofer IME

Figure 29: In vitro depletion assay for pyrene with rainbow trout hepatocytes - Run 2.



Source: own illustration, Fraunhofer IME

Figure 30: In vitro depletion assay for pyrene with rainbow trout hepatocytes - Run 3.



Source: own illustration, Fraunhofer IME

A.6.2 β-HCH

Three runs to test the depletion of β-HCH were performed with rainbow trout hepatocytes. Two runs were performed with cell lot HEP-11-2022 and one additional run was performed using cell lot VK-1 to support the results obtained with lot HEP-11-2022. Both lots had already been used in previous studies and displayed metabolic activities. None of the incubations with active hepatocytes resulted in slopes that are significantly different from zero (Figure 31-Figure 33).

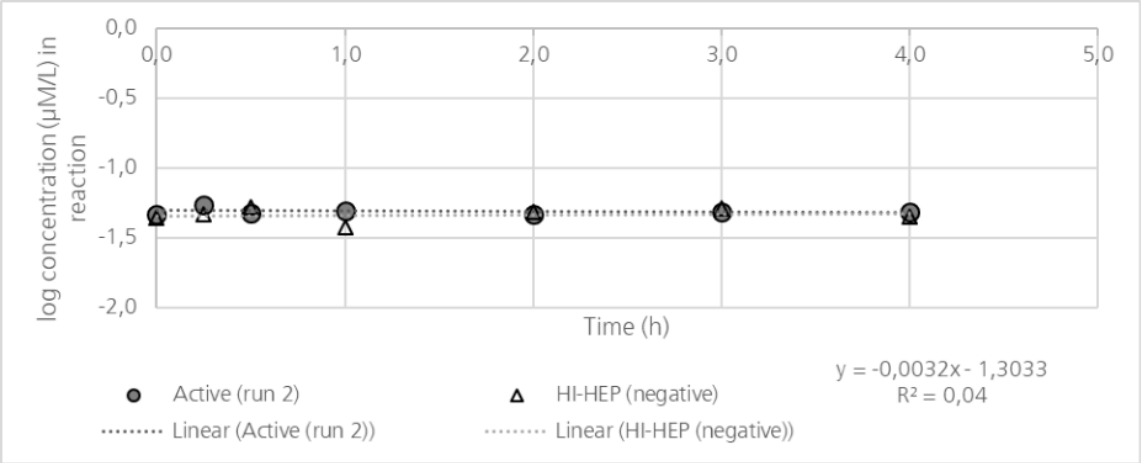
According to the statistical analysis, the HI-slope of run 3 was significantly different from zero, the slope of the active run, however, not.

As Table 41 shows, some validity criteria of OECD TG 319A have not been met. For example, the recovery of one run was below the threshold of 25% and the viability of run 1 with cell lot VK-1 was only 78.6% at test start. Still, since no significant biotic depletion could be detected in any of the runs performed, the data leads to the conclusion that β-HCH is not biotransformed by rainbow trout hepatocytes under the applied test conditions.

Table 41: Summary of the results of the in vitro depletion assays for β-HCH with rainbow trout hepatocytes.

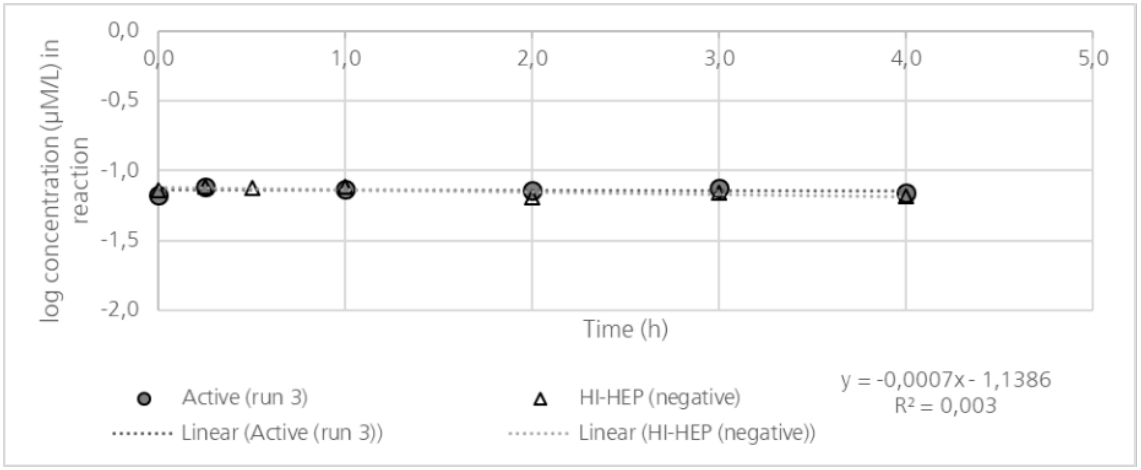
Parameter	Results (data set 1)	Results (data set 2)
Runs	2	1
Cell lot	HEP-11-2022	VK-1
Slope active runs ≠ 0?	No (p = 0.6667 – 0.9194)	No (p = 0.1693)
Runs comparable?	Yes (p = 0.8009)	Only one run performed
HI slopes ≠ 0?	Run 2: No (p = 0.7880) Run 3: Yes (p = 0.0435)	No (p= 0.8582)
HI loss of substance < 20%	Yes	Yes
Cellular concentration (mean ± SD)	2.4 ± 0.04 * 10^6 cells/mL	2.09 ± 0.1 * 10^6 cells/mL
Viability (mean ± SD)	89.9 ± 0.1 %	78.6 ± 2.4%
Recovery (mean ± SD)	24.9 ± 0.9%	41.8%
LOD	0.5 µg/L = 0.0017µM/L	0.5 µg/L = 0.0017µM/L
$CL_{in\ vitro,\ int}$ (mean ± SD)	0 mL/h/10^6 cells	0 mL/h/10^6 cells
Comment	In run 3, one sampling point had to be omitted from the analysis (vial broke).	

Figure 31: In vitro depletion assay for β-HCH with rainbow trout hepatocytes - Run 2.



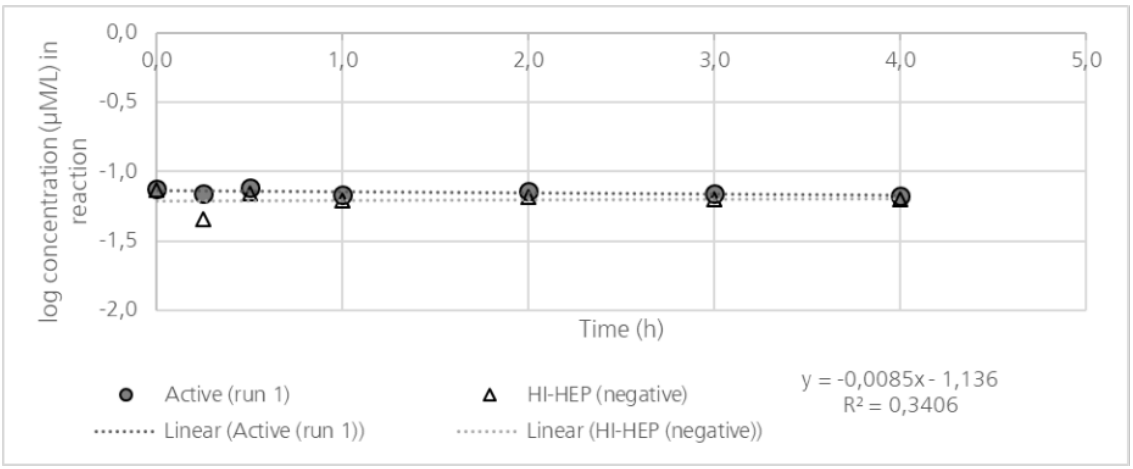
Source: own illustration, Fraunhofer IME

Figure 32: In vitro depletion assay for β-HCH with rainbow trout hepatocytes - Run 3.



Source: own illustration, Fraunhofer IME

Figure 33: In vitro depletion assay for β-HCH with rainbow trout hepatocytes - Run 1.



Source: own illustration, Fraunhofer IME

A.6.3 UV-234

Three runs to test the depletion of UV-234 were performed with rainbow trout hepatocytes (Table 42). Each sample of the active incubation was taken twice and the slope was calculated from the mean concentration of both samplings. For HI-controls only one sample was drawn per sampling time point.

The nominal start concentration of the test substance UV-234 was higher than the reported water solubility of < 5 µg/L at pH 6.1 (*cf*. ECHA Registration Dossier for EC number 274-570-6). It was selected assuming that the solubility may be higher in L-15 medium and taking into consideration that the LOQ was at approx. 4 µg/L. If a depletion had been present, it may not have been measurable. Furthermore, the preliminary tests (*cf*. Chapter A.3.5) indicated that lower concentrations in the test vessel contribute to data scatter, which can reduce the quality of the linear regression.

The statistical analyses show that the slopes that can be calculated from the active incubation data show no significant difference from zero. The same applies to the slopes of the HI incubations, except for run 1. As displayed in Figure 34, the measured concentration slightly increases in the HI incubation of run 1. Since no significant biotic depletion could be detected in any of the runs performed (Figure 34-Figure 36), the data leads to the conclusion that UV-234 is not biotransformed by rainbow trout hepatocytes.

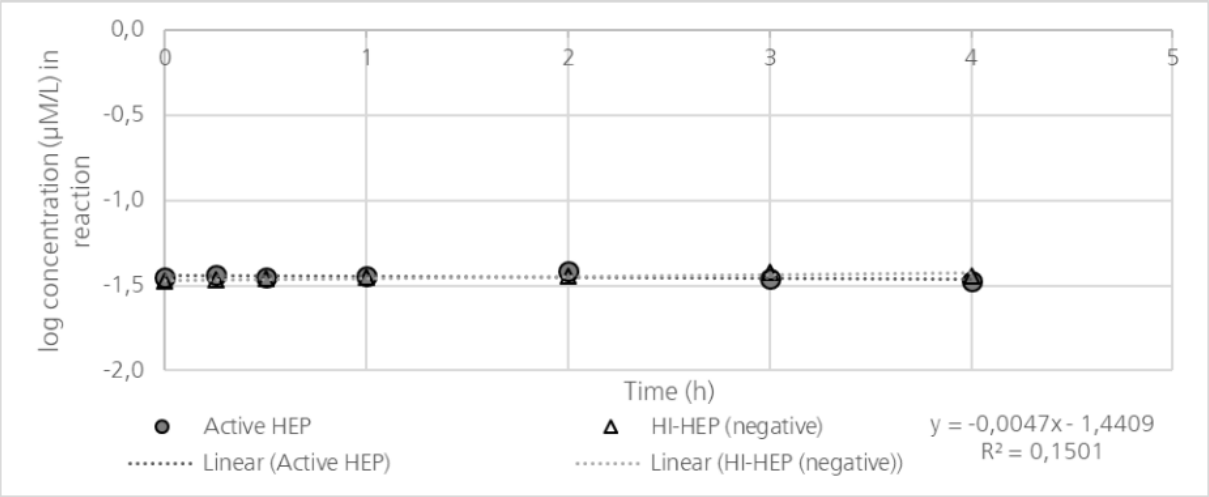
It can be speculated whether a depletion can be detected at lower starting concentrations or by utilizing different dosing techniques. In the literature, three biotransformation products of UV-234 were detected following a six day uptake period of UV-234 with zebrafish embryos. (Zhang et al. 2023). In bioconcentration experiments in zebrafish with six different benzotriazole ultraviolet stabilisers (BUVs) the fish were also analysed for biotransformation products. For only one BUV no biotransformation products were identified and that BUV was UV-234 (Zhang et al. 2021). The literature data indicates that a biotransformation of UV-234 in fish is likely, but that the rate could be very low. It is possible that the applied study design was not suitable to capture low biotransformation rates reliably.

Table 42: Summary of the results of the in vitro depletion assays for UV-234 with rainbow trout hepatocytes.

Parameter	Results
Runs	3
Cell lot	HEP-14-2023
Slope active runs ≠ 0?	No (0.3913 – 0.8899)
Runs comparable?	Yes (p = 0.8983)
HI slopes ≠ 0?	No (p = 0.2414 – 0.6826) Yes (run 1 p = 0.0258)
HI loss of substance < 20%	Yes
Cellular concentration (mean ± SD)	1.89 ± 0.16 * 10^6 cells/mL
Viability (mean ± SD)	81.5 ± 0.7%

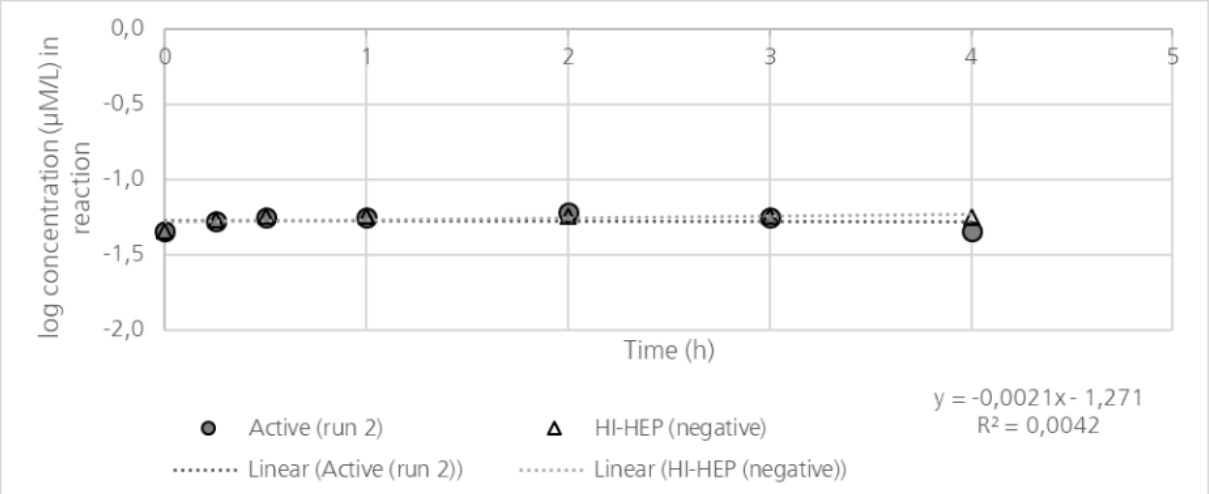
Parameter	Results
Recovery (mean ± SD)	25.8 ± 1.2 %
LOQ	0.0017 µM/L
$CL_{in\ vitro,\ int}$ (mean ± SD)	0 mL/h/10^6 cells
Comment	The slope of the HI-control I run 1 was significantly different from 0, the slope of the active incubation of run 1 was not significantly different from zero

Figure 34: In vitro depletion assay for UV-234 with rainbow trout hepatocytes - Run 1.



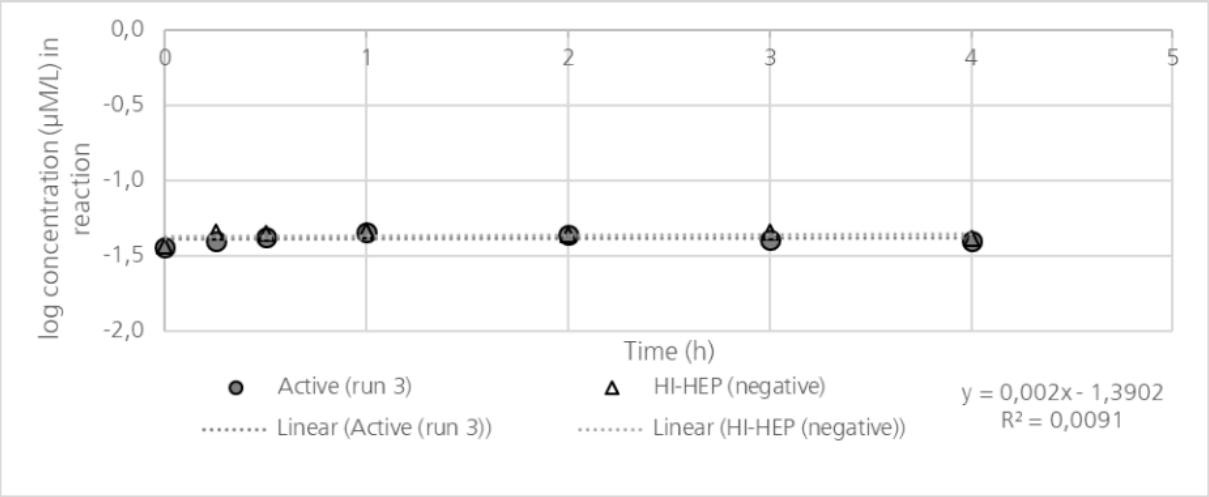
Source: own illustration, Fraunhofer IME

Figure 35: In vitro depletion assay for UV-234 with rainbow trout hepatocytes - Run 2.



Source: own illustration, Fraunhofer IME

Figure 36: In vitro depletion assay for UV-234 with rainbow trout hepatocytes - Run 3.



Source: own illustration, Fraunhofer IME

Note:

Raw data related to the in vitro depletion assays with rainbow trout hepatocytes are available as supplementary material "IVIVE raw data" and "analytics".

A.7 Results – In vitro depletion assays with rat hepatocytes

A.7.1 Pyrene

Three runs to test the depletion of pyrene were performed with rat hepatocytes (Table 43).

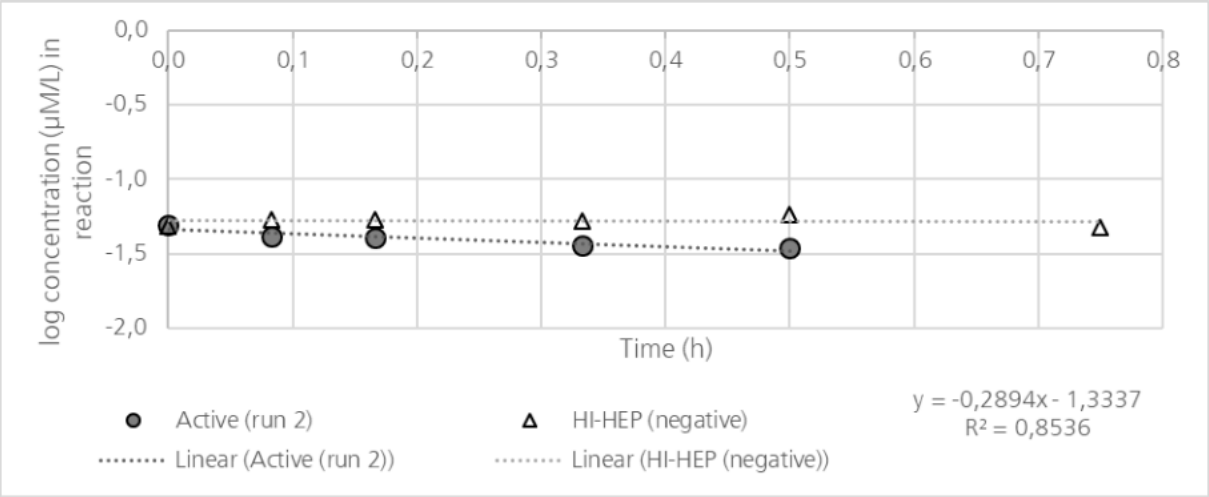
A first run was performed, but only 4 samples of the active incubation could be used for the linear regression. The data of the first run was not used for further evaluations. The slopes of the HI-incubations for runs 3 and 4 are significantly different from zero, but do not show a loss of substance of more than 20%. The slopes of all active incubations are significantly different from zero and from the slope of their HI-controls. The slopes of the active incubations are not significantly different from each other (Figure 37-Figure 39).

Table 43: Summary of the results of the in vitro depletion assays for pyrene with rat hepatocytes.

Parameter	Results
Runs	3
Cell lot	RH191205 Pool (Primacyt)
Slope active runs ≠ 0?	Yes (p = 0.0002279 – 0.02685)
Runs comparable?	Yes (p = 0.6817)
HI slopes ≠ 0?	Yes (run 2 p = 0.8116) No (run 3 p = 0-0260) No (run 4 p = 0.0177)

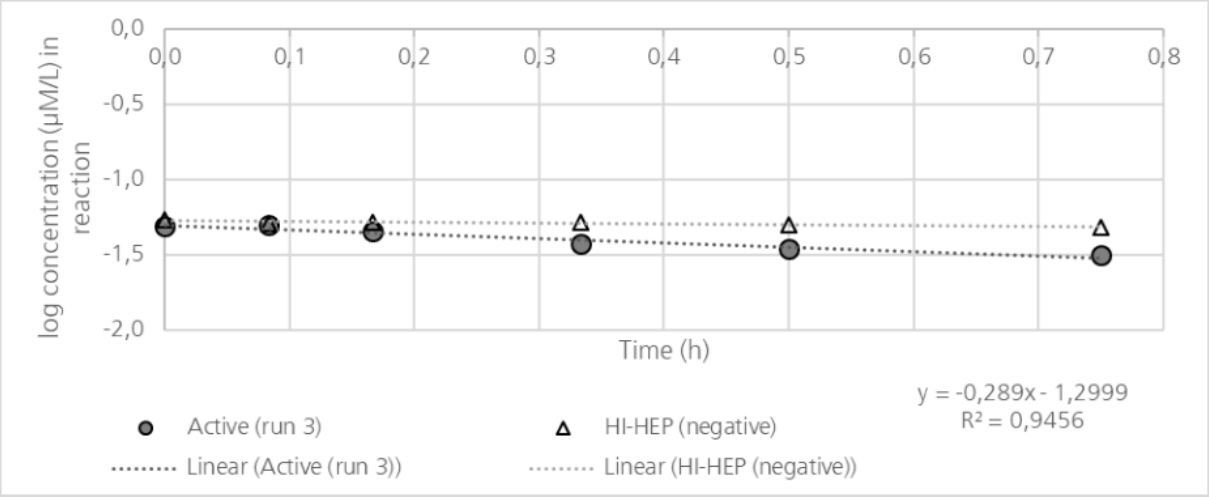
Parameter	Results
HI loss of substance < 20%	Yes
Cellular concentration (mean ± SD)	1.3 ± 0.18 * 10^6 cells/mL
Viability (mean ± SD)	82.2 ± 1.4 %
LOD	0.5µg/L = 0.0247 µM/L
$CL_{in\ vitro,\ int}$ (mean ± SD)	0.55 ± 0.05 mL/h/10^6 cells
Comment	The active incubation of run 2 only consists of 5 datapoints. The datapoint at t = 45 minutes was omitted.

Figure 37: In vitro depletion assay for pyrene with rat hepatocytes - Run 1.



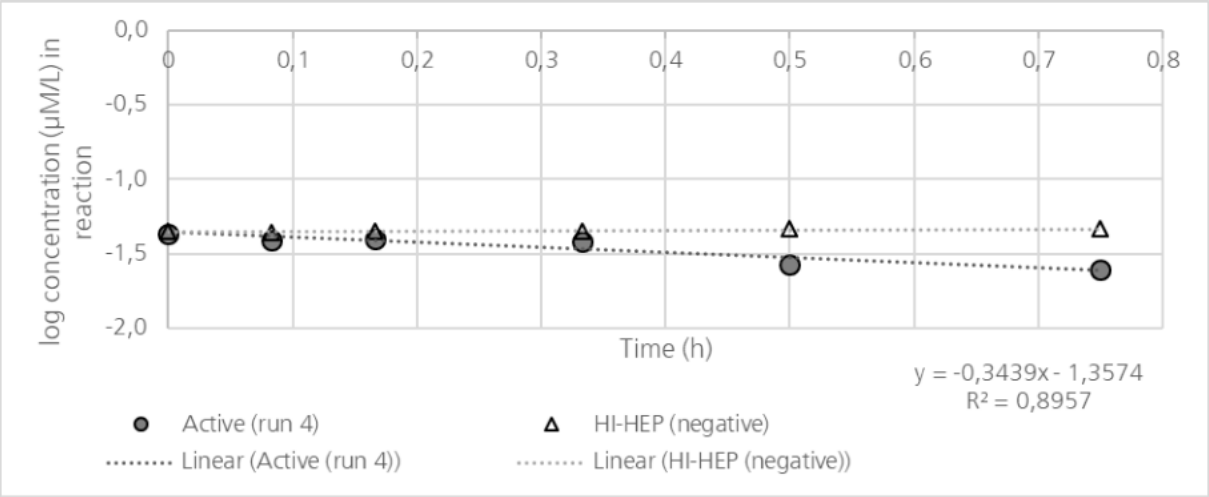
Source: own illustration, Fraunhofer IME

Figure 38: In vitro depletion assay for pyrene with rat hepatocytes - Run 2.



Source: own illustration, Fraunhofer IME

Figure 39: In vitro depletion assay for pyrene with rat hepatocytes - Run 3.



Source: own illustration, Fraunhofer IME

A.7.2 Chlorpyrifos

Two runs to test the depletion of chlorpyrifos were performed with rat hepatocytes. Both runs were performed in parallel, hence also the thawing was performed in parallel (Table 44). The extended time required for the preparation of the cellular material could explain the lower viability that was observed. The viability was still higher than the post-thaw viability reported by the supplier (*cf*. Table 38).

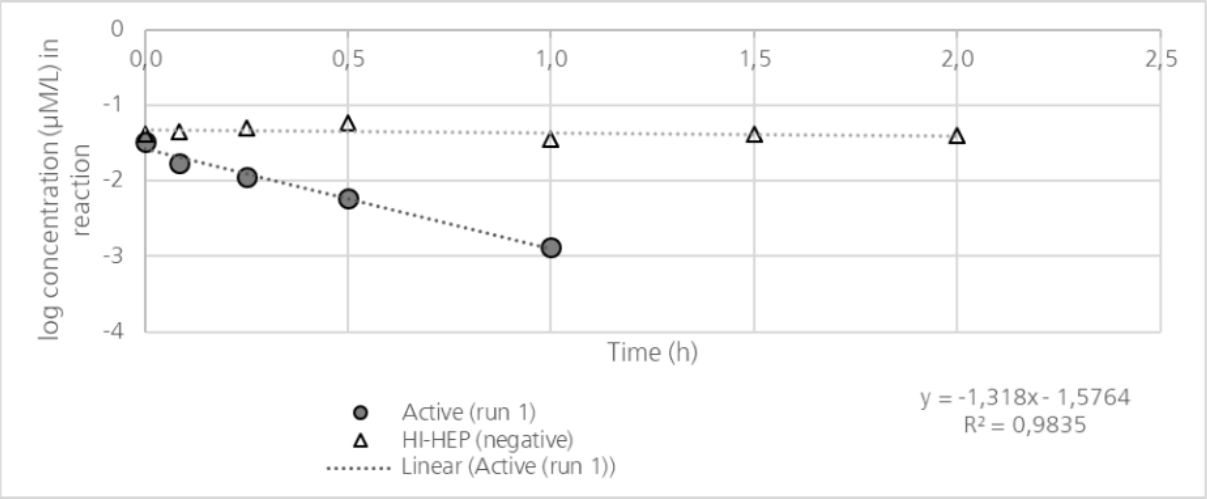
Chlorpyrifos was biotransformed by rat hepatocytes. The incubations were performed for 120 minutes, but after 60 minutes the LOD was reached already, which contradictory to the results of the preliminary experiments. The data of the preliminary tests were quantifiable up to 2hrs of incubation. Due to the low concentrations after 1 hour of incubation in the main tests, only 5 data points were available for the linear regression. The slopes of the active incubations are significantly different from their respective HI-controls but are statistically not different from each other (Figure 40-Figure 41).

Table 44: Summary of the results of the in vitro depletion assays for chlorpyrifos with rat hepatocytes.

Parameter	Results
Runs	2
Cell lot	RH191205 Pool (Primacyt)
Slope active runs ≠ 0?	Yes (p = 0.3157 & 0.3159)
Runs comparable?	Yes (p = 0.4914)
HI slopes ≠ 0?	No
HI loss of substance < 20%	Yes
Cellular concentration (mean ± SD)	0-61 ± 0.0125 * 10^6 cells/mL

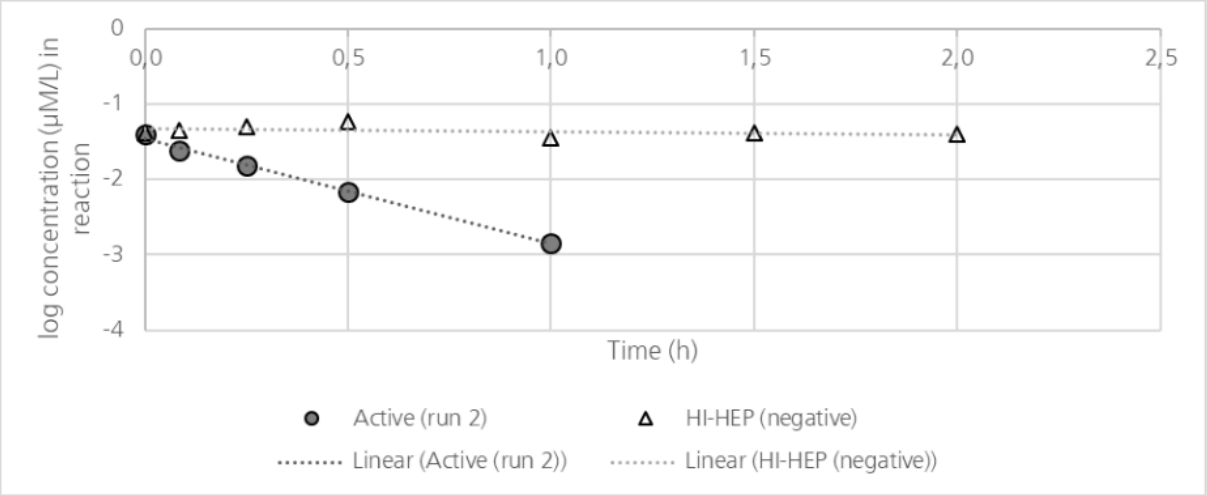
Parameter	Results
Viability (mean ± SD)	77.6 ± 0.4%
LOQ LOD	1.5 µg/L 0.5 µg/L
$CL_{in\ vitro,\ int}$ (mean ± SD)	5.134 ± 0.264 mL/h/10^6 cells
Comment	Runs performed in parallel Only 5 datapoints available for linear regression since LOD was reached at t = 60 minutes.

Figure 40: In vitro depletion assay for chlorpyrifos with rat hepatocytes - Run 1.



Source: own illustration, Fraunhofer IME

Figure 41: In vitro depletion assay for chlorpyrifos with rat hepatocytes - Run 2.



Source: own illustration, Fraunhofer IME

A.7.3 UV-234

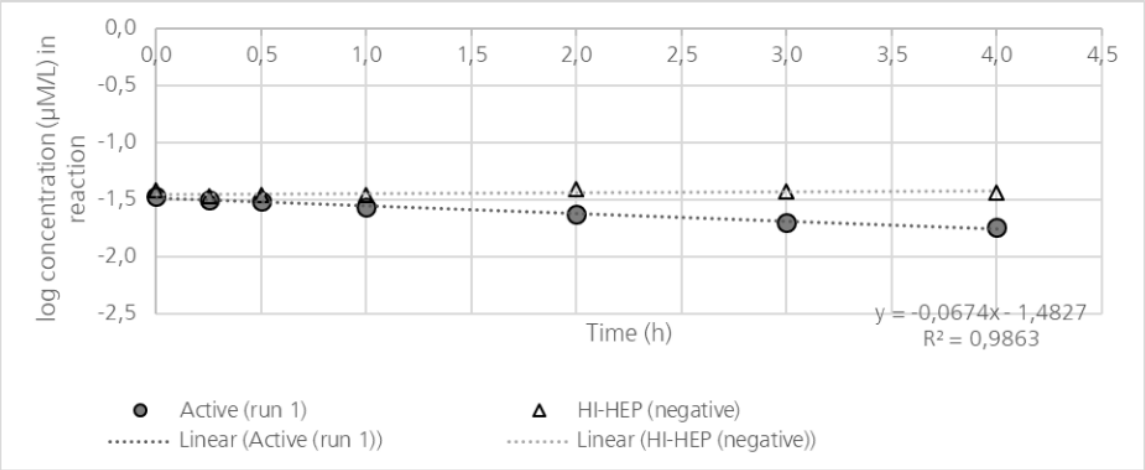
4 runs to test the depletion of UV-234 were performed with rat hepatocytes (Table 45) applying the test setups that are described in Annex A 3.2. In the first run, a depletion of UV-234 was detected in rat hepatocytes (cf. Figure 42). Following this observation, the experimental conditions for the following two runs were adapted by lowering the starting concentrations and by reducing the incubation duration to one hour. As shown in Figure 43 and Figure 44, no statistically significant depletion could be determined under these new conditions. Accordingly, the experimental conditions of the first run were re-applied for the fourth run. No depletion of the test compound could be determined in this run (see Figure 45).

The four runs could not unambiguously resolve whether UV-234 can be biotransformed by rat hepatocytes.

Table 45: Summary of the results of the in vitro depletion assays for UV-234 with rat hepatocytes.

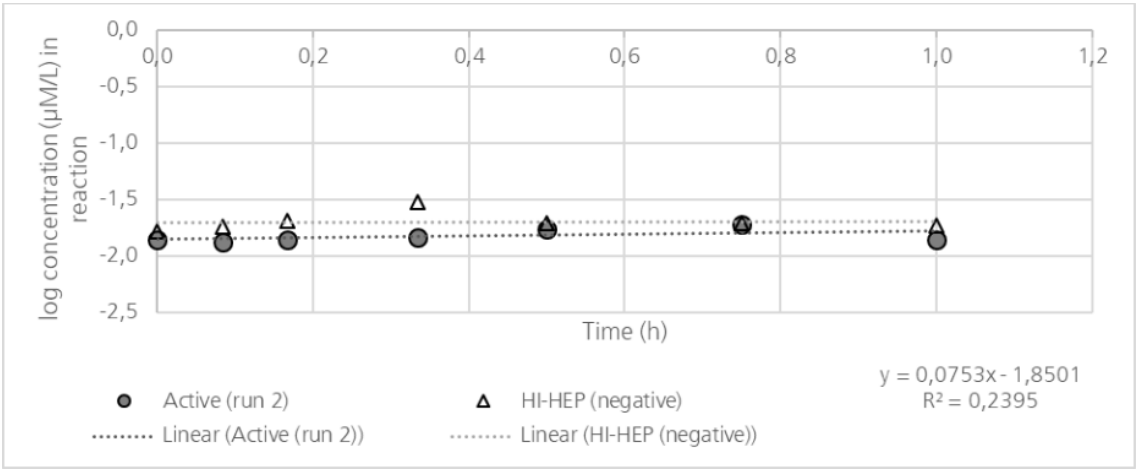
Parameter	Results (data set 1)	Results (data set 2)
Runs	2	2
Cell lot	RH191205 Pool (Primacyt)	RH191205 Pool (Primacyt)
Slope active runs ≠ 0?	Yes (run 1 $p < 0.0001$) No (run 4 $p = 0.4887$)	No ($p = 0.1188$ - 0.2647)
Runs comparable?	No ($p = 0.0194$)	Yes ($p = 0.8220$)
HI slopes ≠ 0?	No (0.2220 – 0.3677)	No ($p = 0.9296$)
HI loss of substance < 20%	Yes	Yes
Cellular concentration (mean ± SD)	Run 1: 1.66 *10^6 cells/mL Run 4: 1.05 *10^6 cells/mL	2.83 ± 0.31 *10^6 cells/mL
Viability (mean ± SD)	Run 1: 82.4% Run 4: 79.6%	83.5 ± 0.2%
LOQ	0.0017 µM/L	0.0017 µM/L
$CL_{in\ vitro,\ int}$ (mean ± SD)	Run 1: 0.09 mL/h/10^6 cells Run 4: 0 mL/h/10^6 cells	0 mL/h/10^6 cells
Comment	Incubation: 4 hours with a nominal start concentration of 0.03 µM	Incubation: 1 hour with a nominal start concentration of 0.01 µM

Figure 42: In vitro depletion assay for UV-234 with rat hepatocytes - Run 1.



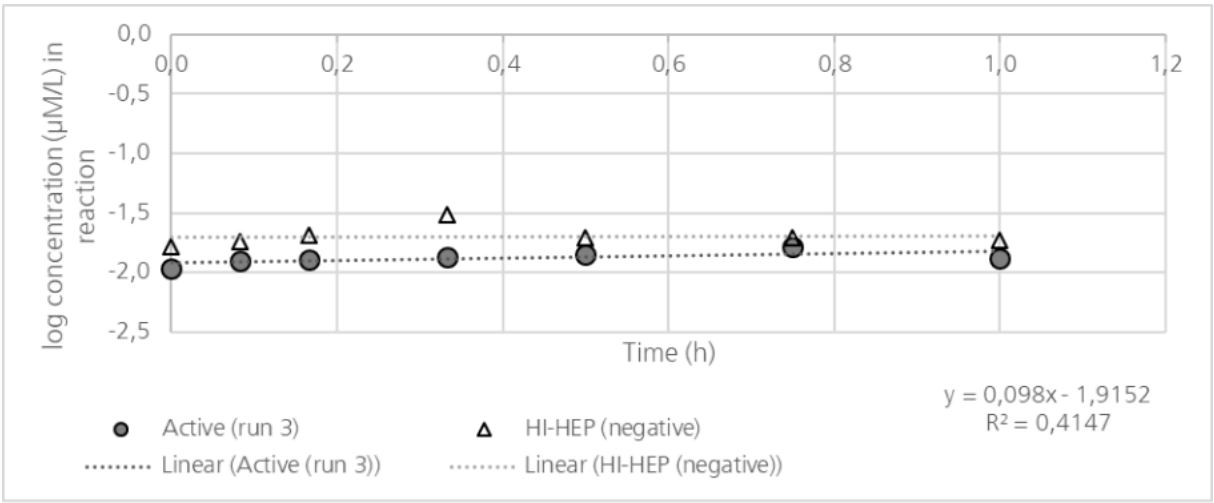
Source: own illustration, Fraunhofer IME

Figure 43: In vitro depletion assay for UV-234 with rat hepatocytes - Run 2.



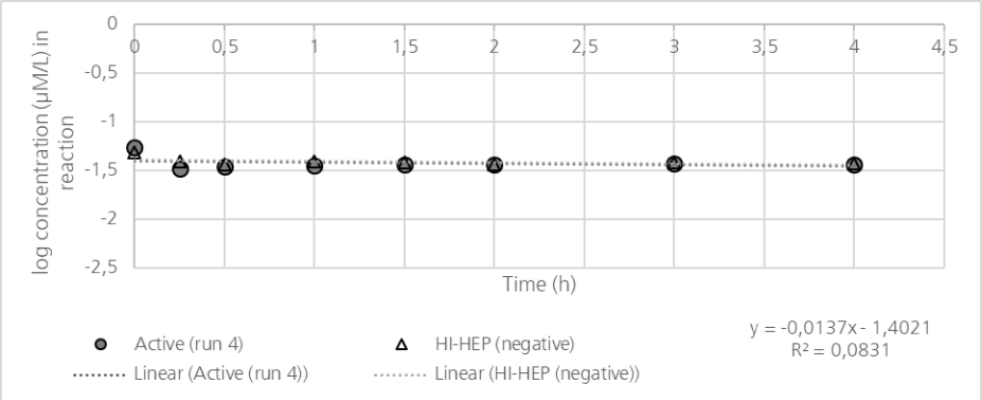
Source: own illustration, Fraunhofer IME

Figure 44: In vitro depletion assay for UV-234 with rat hepatocytes - Run 3.



Source: own illustration, Fraunhofer IME

Figure 45: In vitro depletion assay for UV-234 with rat hepatocytes - Run 4.



Source: own illustration, Fraunhofer IME

Note:

Raw data related to the in vitro depletion assays with rat hepatocytes are available as supplementary material "IVIVE raw data" and "analytics".

A.8 Results – IVIVE BCFs (Rainbow trout)

As described in the material and methods section, the "B-compass Fish" model (Krause & Goss 2020) in version 1.2 was used to extrapolate the in vitro depletion rates to BCF values (Table 46). In addition to the depletion rate data determined in this project, literature depletion rates were also inserted into the model. If more than one study was available, the study that was closest to TG 319A/B criteria was selected.

Table 46: Input parameters and results of the IVIVE BCF calculations for chemicals of the 26 substances set.

Substance	Log *K*ow	Test system	$Cl_{int,\ in\ vitro}$ (mL/h/10^6 cells) (mL/h/mg protein)	Viability (%)	BCF (one-compartment)	Source of depletion rate
Pyrene	4.91	RT HEP	3.59	83.8	144	This project, (A.5.1)
β-HCH	3.87	RT HEP	0	89.9	367	This project, (A.5.2)
UV-234	7.68	RT HEP	0	81.5	7464	This project, (A.5.3)
Anthracene	4.41	RT HEP	0.32	79	647	Fay et al. 2017
Anthracene	4.41	RT-S9	0.49	-	698	Nichols et al. 2013b
Azoxystrobin	3.00	RT HEP	0.59	80.5	34	Kosfeld et al. 2020

Substance	Log *K*ow	Test system	$Cl_{int,\ in\ vitro}$ (mL/h/10^6 cells) (mL/h/mg protein)	Viability (%)	BCF (one-compartment)	Source of depletion rate
B[a]P	6.07	RT HEP	0.54	80	787	Fay et al. 2017
Chlorpyrifos	4.89	RT-S9	1.43	-	509	Zercher et al. 2024
D4	4.98	RT-S9	0.24	-	1935	Cantu et al. 2025
Diclofenac	0.7 (ionised)* 4.35 (neutral)	RT HEP	0.06	82.3	0§ 944§	Kosfeld et al. 2020
Diclofenac	0.7 (ionised)* 4.35 (neutral)	RT-S9	0.15	-	0§ 899§	Pihlaja et al. 2024
17α-ethinylestradiol	3.84	RT-S9	0.09	-	324	EURL ECVAM Database (Halder et al. 2018)
Fluorene	4.06	RT HEP	0.3	79	374	Fay et al. 2017
Fluorene	4.06	RT-S9	0.36	-	433	Nichols et al. 2013b
Fluoxetine	0.93 (ionised)* 4.33 (neutral)	RT HEP	0	86	0§ 1040§	Bourgeois et al. 2022
Fluoxetine	0.93 (ionised)* 4.33 (neutral)	RT-S9	0	-	0§ 1040§	Connors et al. 2013a&b
GenX	3.26	RT HEP	0	N/A	91 §	ECHA Registration Dossier (EC number: 700-242-3) **
Methoxychlor	5.25	RT-S9	0.32	-	1928	Nichols et al. 2018a

Substance	Log *K*ow	Test system	$Cl_{int, in\ vitro}$ (mL/h/10^6 cells) (mL/h/mg protein)	Viability (%)	BCF (one-compartment)	Source of depletion rate
Methoxychlor	5.25	RT HEP	0.08 0.065	80 80	3404 3765	Nichols et al. 2018a Fay et al. 2014
PCB 153	7.08	RT HEP	0	80	13596	Trowell et al. 2018
Pentachlorobenzene	5.11	RT-S9	0	-	5499	Laue et al. 2014
Pentachlorophenol	2.45 (ionised) 4.81 (neutral)	RT-S9	0	-	14§ 2976§	Chen et al. 2016
Phenanthrene	4.43	RT HEP	0.21	79	799	Fay et al. 2017
Phenanthrene	4.43	RT-S9	0.58	-	662	Nichols et al. 2013b
Prochloraz	3.91	RT HEP	0.01	80.5	399	Kosfeld et al. 2020
Simazine	2.12	RT HEP	0	0.8	7	Black et al. 2021
Terbutryn	3.41	RT HEP	1.64	81.3	50	Kosfeld et al. 2020
Triclosan	4.91	RT HEP	3.85	91.5	146	Black et al. 2021
Trifloxystrobin	5.08	RT HEP	0.69	80.5	672	Kosfeld et al. 2020

* Data taken from Köhler et al. 2023

** No details on study available

§ IVIVE model out of applicability domain

A.9 Results – IVIVE BMFs (Rat)

As described in the material and methods section (*cf.* chapter A.4), the Krause and Goss model "B-compass rat" (Krause & Goss 2021) in version 1.0 was used to extrapolate the in vitro depletion rates to BCF values (Table 47). In addition to the depletion rate data determined in this project, literature depletion rates were also inserted into the model.

Compared to the IVIVE BMF values calculated in Lee et al. (2022), the BMF values calculated with the same data and the B-compass rat model are considerably higher.

Table 47: Input parameters and results of the IVIVE BMF calculations for chemicals of the 26 substances set.

Substance	Log K_{OW}	Log K_{OA}	Test system	$Cl_{int, in\ vitro}$ (mL/h/10^6 cells) (mL/h/mg protein)	Viability (%)	BMF (one-compartment)	Source of depletion rate
Pyrene	4.91	8.75	Rat hepatocytes	0.55	82.2	0.773	This project
Chlorpyrifos	4.89	9.47	Rat hepatocytes	5.13	77.6	0.08	This project
UV-234*	7.68	15.32	Rat hepatocytes	0 0.09	83.5 82.4	33 4	This project
B[a]P	6.07	10.98	Rat-S9	0.112	-	3	Lee et al. 2022
β-HCH	3.87	7.59	Rat-S9	0.00103	-	24	Lee et al. 2022
Methoxychlor	5.25	10.62	Rat-S9	0.25	-	1.5	Lee et al. 2022
Pyrene	4.91	8.75	Rat-S9	0.139	-	2.6	Lee et al. 2022
Diclofenac	0.7 (ionised) 4.35 (neutral)	12.86	Rat hepatocytes	3.193 (recalculated from a mean value of 53.21 µL/min/10^6 cells)	80% (assumed)	0[§] 0.12[§]	Louisse et al. 2020

* A significant depletion for this substance could be determined in 1 out of 4 runs. This depletion could not be replicated.

[§] IVIVE model out of applicability domain

A.10 Physicochemical parameters for bioaccumulation assessment: log K_{OW}

The content of this chapter is published as:

Nendza, M., Kosfeld, V., & Schlechtriem, C. (2025). Consolidated octanol/water partition coefficients: Combining multiple estimates from different methods to reduce uncertainties in log K_{ow}. Environmental Sciences Europe, 37, Article 44. https://doi.org/10.1186/s12302-025-01072-2

Parallel investigations were carried out to estimate and consolidate the values of log K_{OW} and log K_{OA}, as well as to identify approaches for obtaining reliable and robust p*K*a estimates. This resulted in sections of the following chapters that correspond to those presented in Nendza et al. (2025).

A.11 Physicochemical parameters for bioaccumulation assessment: log K_{OA}

A.11.1 Estimation of log K_{OA} values and their consolidation

In the assessment of the bioaccumulation potential of organic chemicals, the 1-octanol/air partition coefficient K_{OA} covers the respiratory elimination in air-breathing organisms (Baskaran and Wania 2023). Modelling results from Gobas et al. (2003) indicate that persistent chemicals with log K_{OA} > 5 and log K_{OW} > 2 have the potential to bioaccumulate in terrestrial food chains. In combination with biotransformation half-life < 70 d in humans, Wania et al. (2022) suggested to use log K_{OA} < 5 and log K_{OW} < 2 to identify organic chemicals not subject to bioaccumulation in air-breathing organisms.

Direct measurements of the K_{OA} by generator column or gas chromatography are available for only about 700 substances (Baskaran et al. 2021; Ebert et al. 2023), about twice as many K_{OA} values were derived from experimental K_{OW} and K_{AW} according to the thermodynamic triangle approach:

$$K_{OA} \approx \frac{K_{OW}}{K_{AW}}$$

Estimated log K_{OA} values also use this relationship, but one or both input parameters are calculated instead of measured. This method is included, for example, in EpiSuite (US EPA 2012), ChemProp (UFZ 2021) and EAS-E Suite (ARC 2023). Alternative techniques for log K_{OA} estimation include fragment (group contribution) methods in which the log K_{OA} of a compound is the sum of the incremental contributions of its constituents with constitutional correction factors, e.g. by Ebert et al. (2023). Linear solvation energy relationships (LSER) model the solvation processes, i.e. (1) creation of a cavity in the solvent and (2) incorporation of the solute into the cavity with various solute–solvent interactions (Kamlet et al. 1988) using experimental or calculated descriptors. LSER are implemented, for example, in ChemProp (UFZ 2021), EAS-E Suite (ARC 2023) and UFZ LSERD (Ulrich et al. 2017). An automated read-across model is OPERA (Mansouri et al. 2018), included also in EAS-E Suite (ARC 2023) and the CompTox Chemicals Dashboard (US EPA 2023).

According to their different approaches, the methods for log K_{OA} estimation have different strengths and weaknesses, which may lead to quite different quality of results for different classes of chemicals that may be within or outside the AD of the respective methods. Knowing the accuracy, reliability and variability of log K_{OA} estimates is especially important when substances are assessed against (regulatory) thresholds based on log K_{OA}.

A.11.2 Case study (log K_{OA})

A case study was carried out as part of this project with three main objectives: (1) to estimate multiple log K_{OA} values by different methods for diverse chemicals to exemplify the variability of log K_{OA} values, (2) to systematically analyse the variability of log K_{OA} estimates with regard to the underlying methods and in terms of different chemical classes, and (3) to recommend approaches to obtain reliable and robust log K_{OA} estimates for the hazard and risk assessment of chemicals.

The case study chemicals were selected in the context of the development of an integrated assessment strategy for bioaccumulation in water and air-breathing organisms. Therefore, the

literature was searched for substances with information on bioaccumulation potential from alternative methods, including in vivo, such as the *Hyalella azteca* bioconcentration test (HYBIT) (Schlechtriem et al. 2019; OECD 2024a) and in vitro methods, for example depletion assays with hepatic S9 fractions or hepatocytes, coupled with in vitro-in vivo extrapolation (IVIVE) methods (Nichols et al. 2018a; Kosfeld et al. 2020; Lee et al. 2022). For a broad coverage of applicability domains and to represent relevant chemical classes and influential physicochemical properties, further substances were included in the dataset.

Chemical dataset

The 231 chemicals of this case study represent diverse chemical classes, such as POPs, PCB, PAH, flame retardants, pesticides (insecticides, herbicides, fungicides, …), pharmaceuticals (analgesics, antibiotics, antidepressants, …), fragrances, biocides, UV-filters and plasticisers. Also included are examples of difficult substances such as PFAS, siloxanes and surfactants.

The dataset is chemically diverse, with molecular weights ranging from 54 to 1450 g/mol, mean log K_{OW} from -1.4 to > 10 and mean log K_{OA} from 1.5 to > 15. All collected and calculated data are provided in the supplementary material (log Koa; Table SI-1).

Experimental log K_{OA} data

Measured log K_{OA} values were retrieved from EPISuite (US EPA 2012), the CompTox Chemicals Dashboard (US EPA 2023), EAS-E Suite (ARC 2023) and Ebert et al. (2023). At least one experimental log K_{OA} value was available for 90 chemicals (38 %) of the dataset, ranging from 1.62 to 14.8.

Calculated log K_{OA} data

Estimated log K_{OA} values were obtained by using freely available / public domain tools which are based on different conceptual methods. This selection is exemplary and not exhaustive. The thermodynamic triangle approach ($K_{OA} = K_{OW}/K_{AW}$) was used with EpiSuite (US EOA 2012), ChemProp (UFZ 2021) and EAS-E Suite (ARC 2023). The group contribution methods by Ebert et al. (Ebert et al. 2023) and EAS-E Suite (ARC 2023) were applied. Linear solvation energy relationships (LSER) models were used as available from ChemProp (UFZ 2021), EAS-E Suite (ARC 2023) and UFZ LSERD (Ulrich et al. 2017). Automated read-across (k-nn) was performed with OPERA (Mansouri et al. 2018) as available in EAS-E Suite (ARC 2023) and CompTox Chemicals Dashboard (US EPA 2023). The consensus modelling, pooling two or three estimates, was obtained from EAS-E Suite (ARC 2023). The calculated log K_{OA} data for the case study chemicals, their means across multiple methods, their minimum and maximum values and their ranges, are provided in the supplementary material (log Koa; Table SI-1).

Data analyses

Evaluating the quality of calculated log K_{OA} values requires consistent and reliable reference values. Because for most of the case study chemicals, no measured log K_{OA} data were available, the respective mean of all measured and estimated log K_{OA}, based on up to 21 individual values, was evaluated as an alternative point of reference. The partially available measured data, the respective method means and the complete set of overall mean values were used to compare the variability and range of minimum and maximum estimates across different tools using the same conceptual method and with regard to different chemical classes.

For reasons of consistency, the dataset was split and only 173 substances were used for the quantitative analyses of variability. The remaining 58 substances, i.e. PFAS, siloxanes and surfactants, were regarded separately in order not to complicate the overall picture.

Case study (log K_{OA}): Results

Variability of experimental and calculated log K_{OA} values

To illustrate the variability of log K_{OA} values obtained by different methods, experimental log K_{OA} were collected from different sources (Ebert et al. 2023; US EPA 2012; ARC 2023; US EPA 2023) and estimated by multiple tools based on the thermodynamic triangle approach (US EPA 2012; UFZ 2021; ARC 2023), group contribution methods (Ebert et al. 2023; ARC 2023), LSER models (UFZ 2021; ARC 2023; Ulrich et al. 2017), automated read-across (k-nn) (ARC 2023; US EPA 2023) and consensus modelling (ARC 2023). In total, more than 3000 individual log K_{OA} values were acquired. For most case study chemicals, more than 10 estimates using different methods were obtained. Remaining data gaps are due to the limited scope and applicability domain (AD) of many methods. Complete datasets with 4 experimental and 17 estimated data points are only available for 4 PAH.

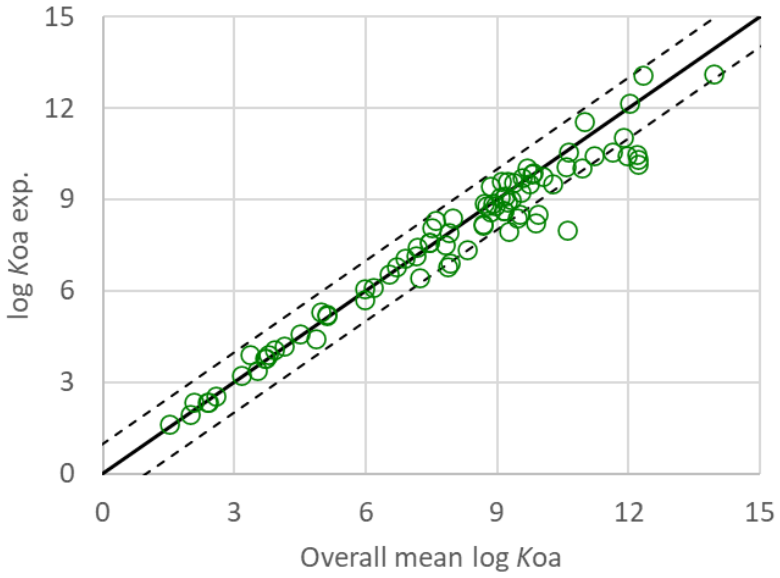
Table 13 summarises the performance of the different methods in terms of availability (number (n) of experimental data reported or estimates obtained), accuracy (percentage of data points within ± 0.5 log units of the overall mean log K_{OA} value (% ± 0.5) and variability represented by the scatter, i.e. largest deviations in either direction (Δ MIN, Δ MAX).

The number of case study chemicals (n) indicates that the availability of experimental data is limited, with less than 40 % even for this set of well-studied substances. In contrast, the availability of data derived by computational methods is much higher and often covers more than 90 % of the case study chemicals. The thermodynamic triangle approach ($K_{OA} = K_{OW}/K_{AW}$), the group contribution methods and the consensus modelling deliver estimates for most chemicals. The different outputs with LSER correspond to the respective availability of either experimental or theoretical descriptors. The applicability of automated read-across depends on the coverage of the underlying experimental database. For data and details see the supplementary material [log Koa; Table SI-2].

The considerable variability of both measured and calculated log K_{OA} is illustrated in Figure 10. Sorted by increasing mean log K_{OA}, the 173 case study chemicals (excluding PFAS, siloxanes and surfactants) reveal log K_{OA} variability mostly within two orders of magnitude. However, with all methods there are also significantly larger deviations (Table 13). For data and details see the supplementary material [log Koa; Table SI-3].

The overall mean log K_{OA} values were used to calculate the performance statistics of the different methods presented in Table 13. The good agreement of the overall mean log K_{OA} with the mean of the available experimental data (r^2 = 0.95, slope = 0.90, intercept = 0.55) is shown in Figure 46. Least scatter is seen with log K_{OA} values below 6. The largest deviations occur at higher log K_{OA} values above 9. If the latter are disregarded, the regression improves significantly (r^2 = 0.97, slope = 0.95, intercept = 0.24) and confirms the adequacy of the overall mean log K_{OA} for assessing the accuracy of log K_{OA} estimates, especially with regard to the log K_{OA} threshold of 5.

Figure 46: Mean of experimental data of 173 case study chemicals compared to the overall mean log K_{OA} across all methods. The solid line is the 1:1 line; the dashed lines indicate deviations of 1 log unit. For data and details see the supplementary material [log K_{OA}; Table SI-4].



Source: own illustration, AL-Luhnstedt

Comparative analyses of variability of log K_{OA} values across methods and chemical classes

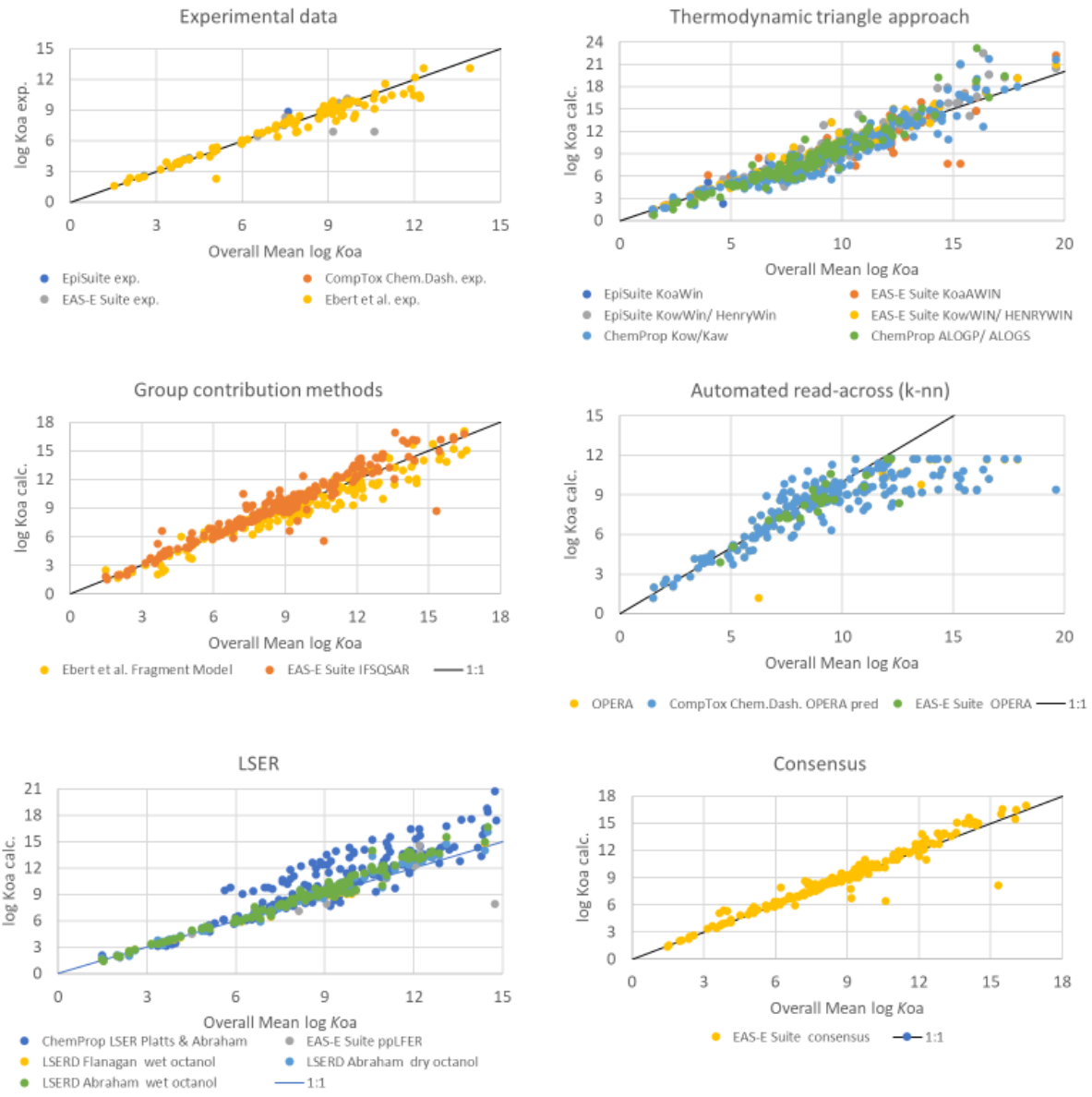
Systematic analyses of the performance of the different approaches for estimating log K_{OA} revealed that the means of the multiple estimates per method (thermodynamic triangle approach ($K_{OA} = K_{OW}/K_{AW}$), group contribution methods, LSER, automated read-across (k-nn), …) correlate with the mean of the experimental data with r^2 about 0.9. This confirms that the different methods are principally suitable and equivalent for estimating the log K_{OA} of the diverse case study chemicals.

The variability of experimental and calculated log K_{OA} obtained by different methods with several tools is shown in Figure 47. The overall mean log K_{OA} values of the case study chemicals were used as consistent point of reference for experimental log K_{OA} from different sources (Ebert et al. 2023; US EPA 2012; ARC 2023; US EPA 2023) and estimated log K_{OA} by the thermodynamic triangle approach (US EPA 2012; UFZ 2021; ARC 2023), group contribution methods (Ebert et al. 2023; ARC 2023), LSER models (UFZ 2021; ARC 2023; Ulrich et al. 2017), automated read-across (k-nn) (ARC 2023; US EPA 2023) and consensus modelling (ARC 2023). There is considerable variability in the experimental and calculated log K_{OA} values. The least scatter is observed for log K_{OA} values below 6. For larger log K_{OA} values, the deviations increase to several orders of magnitude with all the tools. For data and details see the supplementary material [log Koa; Table SI-5].

Variability of log K_{OA} within different classes of chemicals is presented in Figure 48. The POPs, PCB, PAH and flame retardants have largely consistent log K_{OA} values. The least variability can be seen for small molecules with simple chemistry (volatiles, log $K_{OA} < 5$). The pesticides and the fragrances show a uniform variability of about 2 log units. The log K_{OA} values of the pharmaceuticals are even more variable. Their diverse structures and ionisation potential may

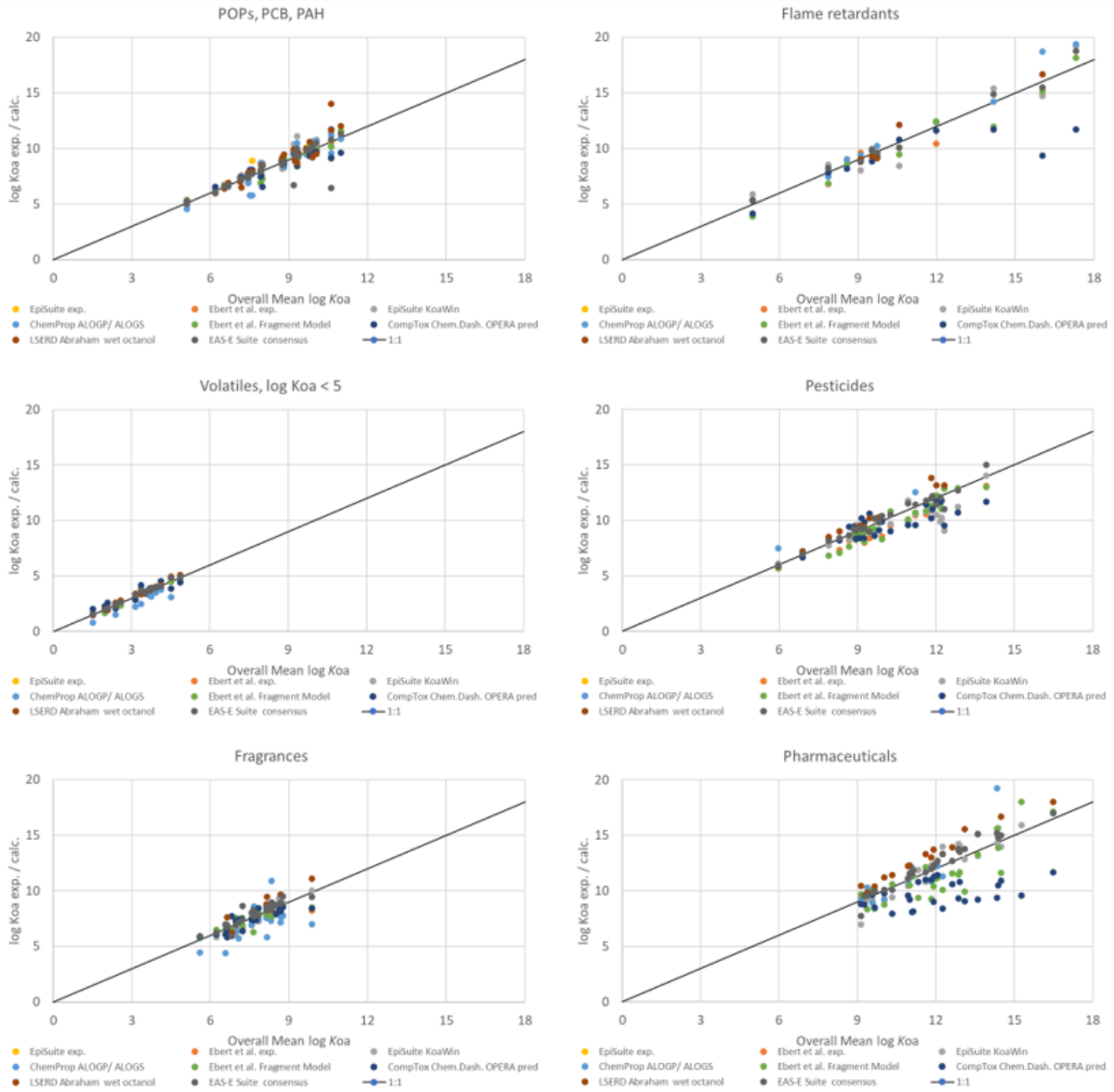
be important factors contributing to the scatter. For data and details see the supplementary material [log K_{OA}; Table SI-6].

Figure 47: **Comparative analysis of the variability of log K_{OA} values of up to 173 case study chemicals obtained by different methods with several tools. The solid line is the 1:1 line. Point of reference are, on the x-axes, the overall mean log K_{OA} across all methods. See Table 13 for details on the performance of the tools and references.**



Source: own illustration, AL-Luhnstedt

Figure 48: **Comparative analysis of the variability of log K_{OA} values of different classes of case study chemicals. The solid line is the 1:1 line. Point of reference are, on the x-axes, the overall mean log K_{OA} across all methods. See Table 13 for details on the performance of the tools and references.**

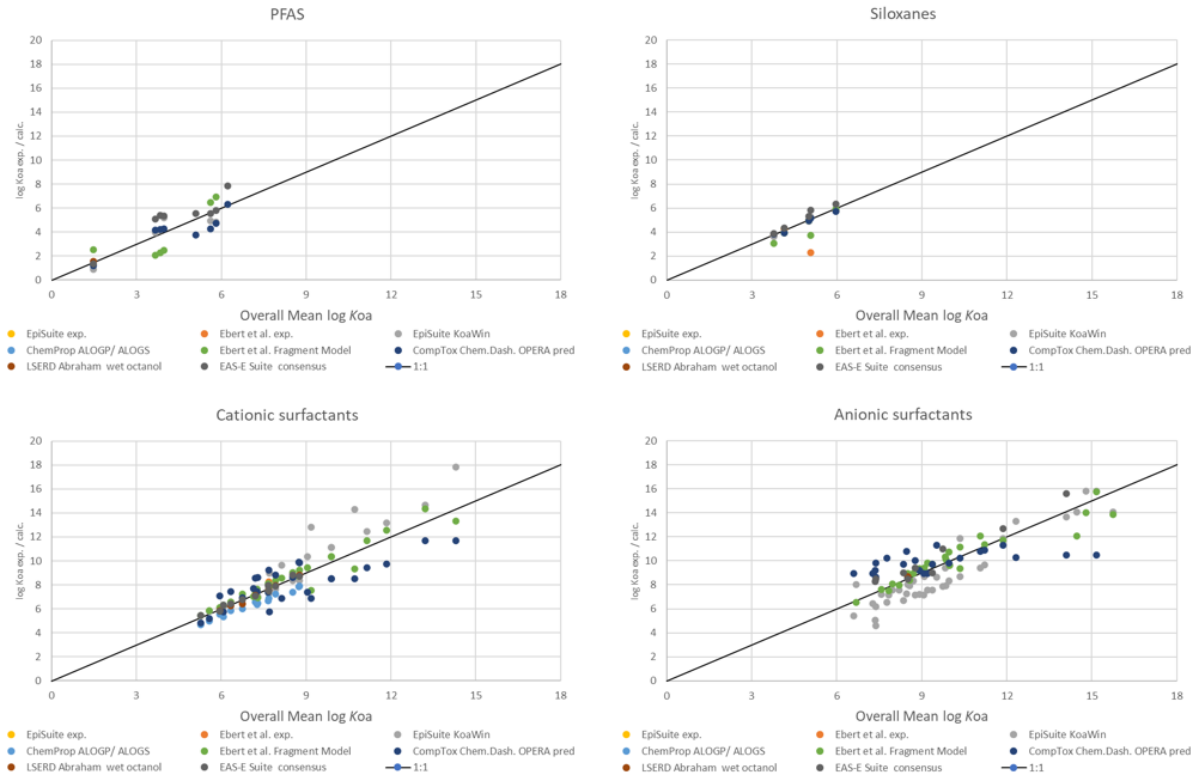


Source: own illustration, AL-Luhnstedt

Difficult substances: PFAS, siloxanes and surfactants

For some compound classes, where the determination of log K_{OW} can be difficult or even impossible, also separate analyses of the log K_{OA} estimates are presented (Figure 49). The PFAS and siloxanes among the case study chemicals are mostly volatile with log K_{OA} values below 7, yet the estimates vary by 2 or more orders of magnitude. The estimated log K_{OA} of the anionic and cationic surfactants among the case study chemicals are all above 5, thus not relevant in the context of terrestrial bioaccumulation assessment. Further analyses of the scatter are not meaningful; the datasets are too small to reveal statistically significant trends.

Figure 49: **Comparative analysis of the variability of log K_{OA} values of difficult substances among the case study chemicals (PFAS, siloxanes and surfactants) obtained by different methods with several tools. The solid line is the 1:1 line. Point of reference are, on the x-axes, the overall mean log K_{OA} across all methods. See Table 13 for details on the performance of the tools and references. For data and details see the supplementary material [log K_{OA}; Table SI-6].**



Source: own illustration, AL-Luhnstedt

Case study (log K_{OA}): Discussion

As with log K_{OW}, the measured and calculated log K_{OA} values show major variability of more than one log unit. Consensus modelling, i.e. a WoE or averaging approach, is therefore recommended to reduce uncertainties and consolidate variable, possibly conflicting, results. This applies whenever the entire range of log K_{OA} values is considered. However, if the question is whether a log K_{OA} value is below or above the threshold of 5, as in the case of bioaccumulation assessment, a reduced approach is sufficient.

Consensus modelling: consolidated log K_{OA} values

Consolidation of log K_{OA} values by consensus modelling reduces uncertainties and is superior to the traditional preference for a single measured log K_{OA}. Combining multiple estimates from different methods takes advantage of their particular strength while compensating outliers. Even with the reduced approach, a single value is too uncertain for a robust decision on log K_{OA} less or greater than 5. Here, at least 2 log K_{OA} values are combined. These can be validated experimental data and estimates calculated by different conceptual approaches, such as the thermodynamic triangle approach ($K_{OA} = K_{OW}/K_{AW}$), a group contribution (atom/fragment) method, a linear solvation energy relationship (LSER), and automated read-across (k-nn). Estimates from other methods with alternative approaches such as quantum-chemistry-based and deep learning neural networks are also suitable. Provided that the computational methods for log K_{OA} estimation fulfil the OECD criteria for scientific validity (OECD 2007) and that the

target compounds fit within their AD, reliable consolidated log K_{OA} with reasonable uncertainties can be obtained. The OECD QSAR Assessment Framework is a comprehensive documentation tool for the assessment of the reliability of QSAR predictions in regulatory contexts (OECD 2023).

Box A1 outlines a tiered procedure including the reduced WoE or averaging approach that combines estimates from at least 2 different methods. The crucial first step is the verification of the chemical structures and their proper description, mostly by SMILES, with due regard to tautomers and speciation. This is usually more time-consuming than the subsequent input to several freely available / public domain, commercial or in-house software to calculate log K_{OA}. If the model output includes information about reliability of estimates, such as AD coverage, confidence intervals and results for similar substances, highly rated estimates are preferred. Erroneous and uncertain estimates, that may be due to AD violation, and obvious outliers should be discarded. Indicators of the quality of estimates, such as reliability ratings and similarity indices, may further be input to weight the evidence.

Box A1. Consensus modelling for consolidated log K_{OA} values.

Tiered procedure:

(1) Verify chemical identity (2D and 3D structure, not only CAS), with due regard to tautomers and speciation.

(2) Check AD compliance, considering specific AD limitations of individual models.

(3) Calculate/determine 2 log K_{OA} values (non-ionised chemical) with different independent methods, including, e.g., quality-controlled experimental data and results from the thermodynamic triangle approach ($K_{OA} = K_{OW}/K_{AW}$), group contribution (atom/fragment) methods, linear solvation energy relationships (LSER), automated read-across (k-nn), quantum-chemistry, or deep learning neural networks.

(4) Analyse results for outliers, eliminate obvious misestimates.

(5) If the 2 log K_{OA} values are within 1 log unit and less than 4, apply WoE or averaging approach → **consolidated log K_{OA}.**

(6) If the 2 log K_{OA} values are between 4 and 6, repeat steps (3) and (4) using different independent methods until a robust conclusion is feasible: apply WoE or averaging approach → **consolidated log K_{OA}.**

(7) If the 2 log K_{OA} values are above 5, the calculations end here if a precise log K_{OA} above 5 is not needed. Else, repeat steps (3) and (4) using different independent methods until at least 4 values within 1 log unit are obtained: apply WoE or averaging approach → **consolidated log K_{OA}.**

The reduced approach takes advantage of the fact that the margin of error in log K_{OA} values below 7 appears limited to 1 log unit (Figure 11). Thus, if the available log K_{OA} values are all within 1 log unit and less than 4, it is sufficiently likely that their (weighted) mean is below the threshold of 5 to identify organic chemicals with low potential for bioaccumulation in air-breathing organisms. Estimates between 4 and 6 may require more data to come to a robust conclusion. If the estimates of log K_{OA} for a substance are above 5, the (weighted) mean of 4 or more reliable experimental and estimated data is the consolidated log K_{OA}.

The performance of the consensus modelling is demonstrated with the case study chemicals (Figure 11), using freely available / public domain tools: EPISuite KoaWin (US EPA 2012) and ChemProp ALOGP/ALOGS (UFZ 2021) for the thermodynamic triangle approach (K_{OA} = K_{OW}/K_{AW}), the group contribution method by Ebert et al. (2023), the Abraham (wet octanol) model in UFZ LSERD (Ulrich et al. 2017) and the consensus log K_{OA} of EAS-E Suite (ARC 2023). The mean of their results correlates well with the overall mean log K_{OA} values across all methods (r^2 = 0.98, slope = 1.01, intercept = 0.01). The data, statistics and figures are provided in the supplementary material (log Koa; Table SI-7).

Conclusions

Three key findings emerge from the above analyses: (1) measured and calculated log K_{OA} reveal considerable variability up to one log unit and more, (2) no method (experimental or computational) is consistently superior for determining log K_{OA}, and (3) the consolidated log K_{OA}, i.e. the (weighted) average of all valid measured and estimated data per chemical, is robust and highly reproducible, scientifically credible, as well as cost efficient.

The limited variability of low log K_{OA} (within one order of magnitude) allows a reduced approach to assess substances against the threshold of 5 for unlikely bioaccumulation in air-breathing organisms.

Acknowledgments

Thanks! to Nadin Ulrich and Ralf-Uwe Ebert (UFZ, Leipzig, Germany) for contributing log K_{OA} estimates.

A.12 Physicochemical parameters for bioaccumulation assessment: p*K*a

A.12.1 Acid dissociation constants: Data availability, estimation methods and uncertainties of p*K*a

The extent to which an organic acid or base bioaccumulates in aquatic organisms depends primarily on two physicochemical properties, the 1-octanol/water partition coefficient log K_{OW} and the acid dissociation constant p*K*a.

The fractions of the species of a compound (neutral, anionic, cationic, zwitterionic) depend on the pH value of the surrounding phase. Evidently, the deprotonated (anionic) form of an acid has a different polarity than the non-dissociated form, which results in reduced hydrophobicity and increased water solubility. As a result, the bioaccumulation of this compound may decrease as compared to the neutral form. Only in few cases does the opposite effect occur; for example, when ionic compounds interact with charged membrane components acting as an ion trap (Nendza 1998).

The influence of different pH values on the speciation of a compound leads to more complex distribution phenomena (Schwarzenbach et al. 1993; Nendza 1998; Nendza and Hermens 1995; Schüürmann et al. 2007; Köhler et al. 2023). For acids AH and bases B that can donate or accept one proton, the unionised and ionised fractions fu and fi depend on p*K*a and pH according to the Henderson-Hasselbalch relationship:

$$f_{u\,acid} = f_{i\,base} = \frac{1}{1+10^{\,pH-pKa}}. \qquad \text{Eq. 8}$$

$$f_{i\ acid} = f_{u\ base} = \frac{1}{1+10^{\ pKa-pH}} \quad \text{Eq. 9}$$

Equation 8 quantifies the undissociated compound fraction AH of the acid as well as the protonated (ionised) form BH^+ of base B. In the latter case, p*K*a(H) refers to BH^+ as the conjugate acid of B. Correspondingly, Equation 9 yields the fraction of the dissociated (ionised) form A^- of AH as well as the unprotonated (neutral) form of B (Schüürmann et al. 2007)

The observed distribution coefficient of acids and bases (D_{OW}) is given by Equation 10, with K_{OW} denoting the partition coefficient of the neutral compound and *K*i the partition coefficient of the ionised compound (A- or BH+).

$$D_{OW} = f_u \cdot K_{OW} + f_i \cdot Ki \quad \text{Eq. 10}$$

Generally, the partitioning of ionised species into other than aqueous phases by passive diffusion is highly unlikely. Therefore, Equation 10 can be simplified and the observed distribution is related to the log K_{OW} values of the non-dissociated form by the following equations for acids (Equation 11) and bases (Equation 12), respectively:

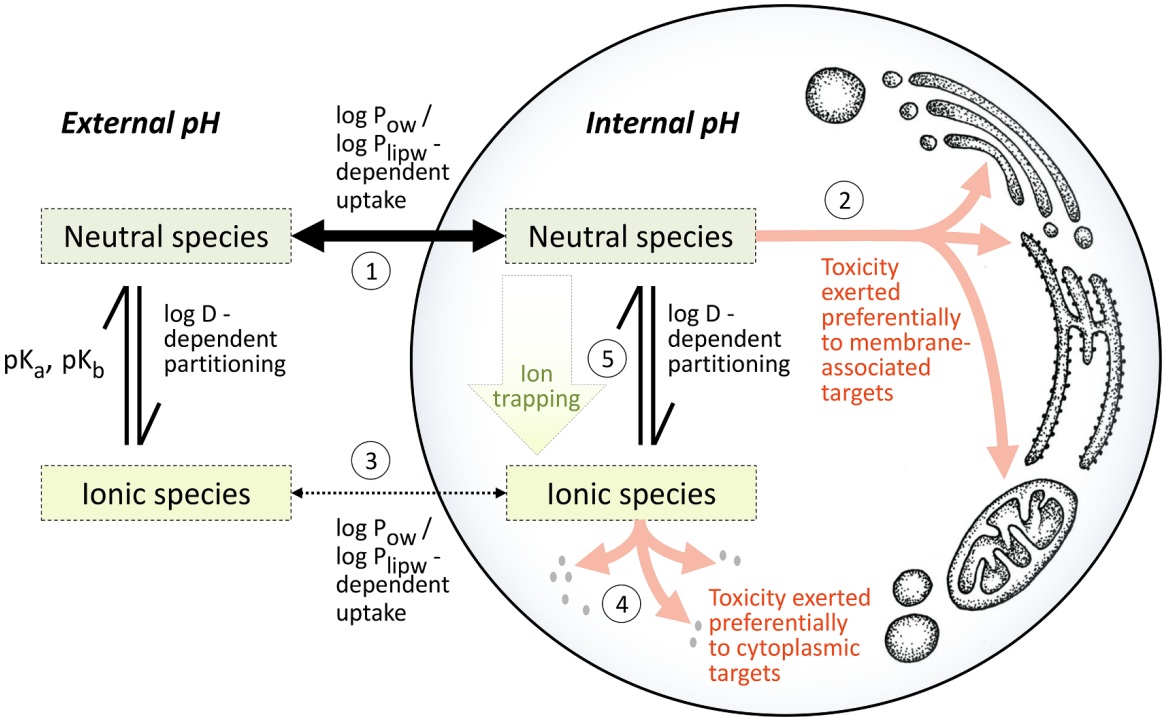
$$\log D_{OW} = \log K_{OW} - \log (1 + 10^{\,pH - pKa}) \text{ (acids)} \quad \text{Eq. 11}$$

$$\log D_{OW} = \log K_{OW} - \log (1 + 10^{\,pKa - pH}) \text{ (bases)} \quad \text{Eq. 12}$$

The above terms are applicable when the concentrations of the ionised species in the organic phase are negligible, and the ambient pH is relatively close to the respective p*K*a values.

The pH partition profiles take the form of a sigmoidal curve. The log D_{OW} corresponds to the log K_{OW} of the unionised compound at pH values < p*K*a for acids or > p*K*a for bases (log D_{OW} = log K_{OW}). Increased ionisation due to pH changes results in a linear decrease in the log D_{OW} until another plateau is achieved where only the ionised form determines the partitioning. The situation becomes much more complex if the molecule contains multiple dissociable groups or if ion pair formation has to be accounted for (Nendza 1998). Figure 50 illustrates the distribution of charged (ionic) and uncharged (neutral) species of ionisable organic chemicals and their possible uptake pathways into cells (Köhler et al. 2023).

Figure 50: **Distribution of charged (ionic) and uncharged (neutral) species of ionisable organic chemicals, their possible uptake pathways into the cell (1, 3) and possible targets of toxic action (2, 4). A pH-specific distribution of charged and uncharged species exists both outside and inside the cell. It can be assumed that the diffusion and thus the uptake of the uncharged species through the outer cell membrane (1) is easier and therefore more relevant than the uptake of the charged species (3). The intracellular conversion of uncharged molecules to charged, poorly membrane-permeable molecules leads to the establishment of an equilibrium (5) corresponding to the cytoplasmic pH and is called ion trapping. Reproduced from Köhler et al. (2023).**



Under environmental conditions, i.e. water bodies having pH between 5 and 9, only weak to moderate acids (p*K*a > 4) and bases (p*K*a(H) < 10) can be partially ionised with pH-dependent partitioning profiles (Table 14). The bioaccumulation of strong acids (p*K*a < 4) and bases (p*K*a(H) > 10) is not influenced by the ambient pH, as it is only due to the ionised species of the compounds. Strong acids and bases are therefore not considered here.

Experimental methods for determining p*K*a are well established. The most common procedures for p*K*a measurements include titration, conductometry, electrometry, photometry and magnetic resonance. The p*K*a values are commonly measured at 20°C or 25°C, and at a given ionic strength (e.g. 0.05-0.1 M salt solution). Compilations of p*K*a values are available, for example (Perrin et al. 1981; Slater et al. 2014; OECD 2023; ECHA 2023). The p*K*a values reported in the literature for a given organic acid or base may vary by 0.3 p*K*a units, depending on the type of measurement and the conditions chosen (Schwarzenbach et al. 1993). Practical examples reveal even larger ranges, e.g., in the ECHA registration dossier of ethylparaben (CAS 120-47-8), the key value of 8.4 was obtained by considering several reliable values in the range of 8.18 to 8.9 in a weight-of-evidence approach.[2]

[2] https://echa.europa.eu/registration-dossier/-/registered-dossier/13843/4/22 (accessed 09/04/2024).

Computational methods for estimating p*K*a values are often based on linear free energy relationships (LFER), using Hammet-type equations for chemical classes sharing the same ionisable functional groups, such as phenols, benzoic acids, anilines, substituted pyridines, etc. (Perrin et al. 1981; Schwarzenbach et al. 1993; Nendza 1998; Nendza and Hermens 1995; Lyman et al. 1982; ACD/Labs 2023).

$$pKa = pKa' - \rho (\Sigma \sigma) \quad \text{Eq. 13}$$

The p*K*a value of the substituted derivatives can be calculated from that of the parent compound (p*K*a '), the substance class-specific constant ρ, which is a measure of the sensitivity of the reaction to substituent effects with respect to the ionisation of benzoic acids *(*ρ = 1), and tabulated σ-Hammett values (Nendza 1998) (Equation 13). An alternative are read-across approaches using endpoint information for one or more chemicals (source analogues) to predict the same endpoint for another substance (target) that is considered structurally similar (ECHA 2017c). Automated read-across, e.g. (Mansouri et al. 2018; US EPA 2023), uses a number of close analogues (k-nearest neighbours (k-nn)) with a minimum chemical similarity distance. Performance depends on the quality of the source data; reliable predictions are only possible if the chemical class in question is adequately represented in the database. Further options for p*K*a calculations are SPARC (ARChem 2010), using computational algorithms based on fundamental chemical structure theory to estimate a wide variety of reactivity parameters strictly from molecular structure, and Chemicalize (Chemaxon 2024), available for small molecules (≤ 12 non-hydrogen atoms) with ChemSpider (Royal Society of Chemistry 2018). Machine learning methods for p*K*a prediction have been described by, e.g., (Mansouri et al. 2019; Navo & Jiménez-Osés. 2021; Luo et al. 2024). While there are quite a number of proprietary software packages for the prediction of p*K*a, the availability of free and open-source programmes for this purpose is limited.

A.12.2 Case study (p*K*a)

A case study was carried out as part of this study having three main objectives: (1) to obtain several p*K*a values by different methods for the case study chemicals to exemplify the variability of p*K*a values, (2) to analyse the variability of p*K*a estimates regarding the underlying methods and the different chemical classes, and (3) to recommend approaches to obtain reliable and robust p*K*a estimates for the hazard and risk assessment of chemicals.

The case study chemicals were selected in the context of the development of an integrated assessment strategy for bioaccumulation in water and air-breathing organisms. Therefore, the literature was searched for substances with information on bioaccumulation potential from alternative methods, including in vivo, such as the *Hyalella azteca* bioconcentration test (HYBIT) (Schlechtriem et al. 2019; OECD 2024) and in vitro methods, for example depletion assays with hepatic S9 fractions or hepatocytes, coupled with in vitro-in vivo extrapolation (IVIVE) methods (Nichols et al. 2018a; Kosfeld et al. 2020; Lee et al. 2022). For a broad coverage of applicability domains and to represent relevant chemical classes and influential physicochemical properties, further substances were included in the dataset.

Chemical dataset

The 231 chemicals selected for this case study represent diverse chemical classes, such as POPs, PCB, PAH, flame retardants, pesticides (insecticides, herbicides, fungicides, …), pharmaceuticals (analgesics, antibiotics, antidepressants, …), fragrances, biocides, UV-filters and plasticisers. Also included are examples of PFAS, siloxanes and surfactants.

The dataset is chemically diverse, with molecular weights ranging from 54 to 1450 g/mol, mean log K_{OW} from -1.4 to > 10, and mean log K_{OA} from 1.5 to > 15. Among the 231 case study chemicals are 25 strong acids (p*K*a < 4) and 9 strong bases (p*K*a(H) > 10). 33 acids and 26 bases have p*K*a or p*K*a(H) between 4 and 10, respectively. Their collected and calculated data are provided in the supplementary material (pKa; Tables SI-1, SI-2, SI-3).

Experimental *pKa* data

A total of 63 p*K*a values were retrieved from ECHA (2023), the CompTox Chemicals Dashboard (US EPA 2023) and EAS-E Suite (ARC 2023). About one third of these, however, are not measured but estimated data. Furthermore, some of the reported values in the ECHA database (ECHA 2023) are obviously *K*a values and not p*K*a values, these were transformed accordingly for this study (pKa; Table SI-1).

Other sources of experimental information, e.g. the IUPAC repositories (Slater 2014), OECD (2023) eChemPortal, OECD (2023a) QSAR Toolbox, were not considered for this study.

Calculated p*K*a data

Estimated p*K*a values were obtained by two different conceptual methods. **LFER (Hammet equation)-based methods** were used with ACD/Labs (2023). **Automated read-across** (k-nn) was obtained with OPERA (Mansouri et al. 2018) and the CompTox Chemicals Dashboard (USEPA 2023). Furthermore, the degree of ionisation as provided by ACD/Labs (2023), OPERA (Mansouri et al. 2018) and EAS-E Suite (ARC 2023) was evaluated.

The calculated p*K*a data for the case study chemicals, their means and their ranges, are provided in the supplementary material (pKa; Tables SI-1, SI-2, SI-3).

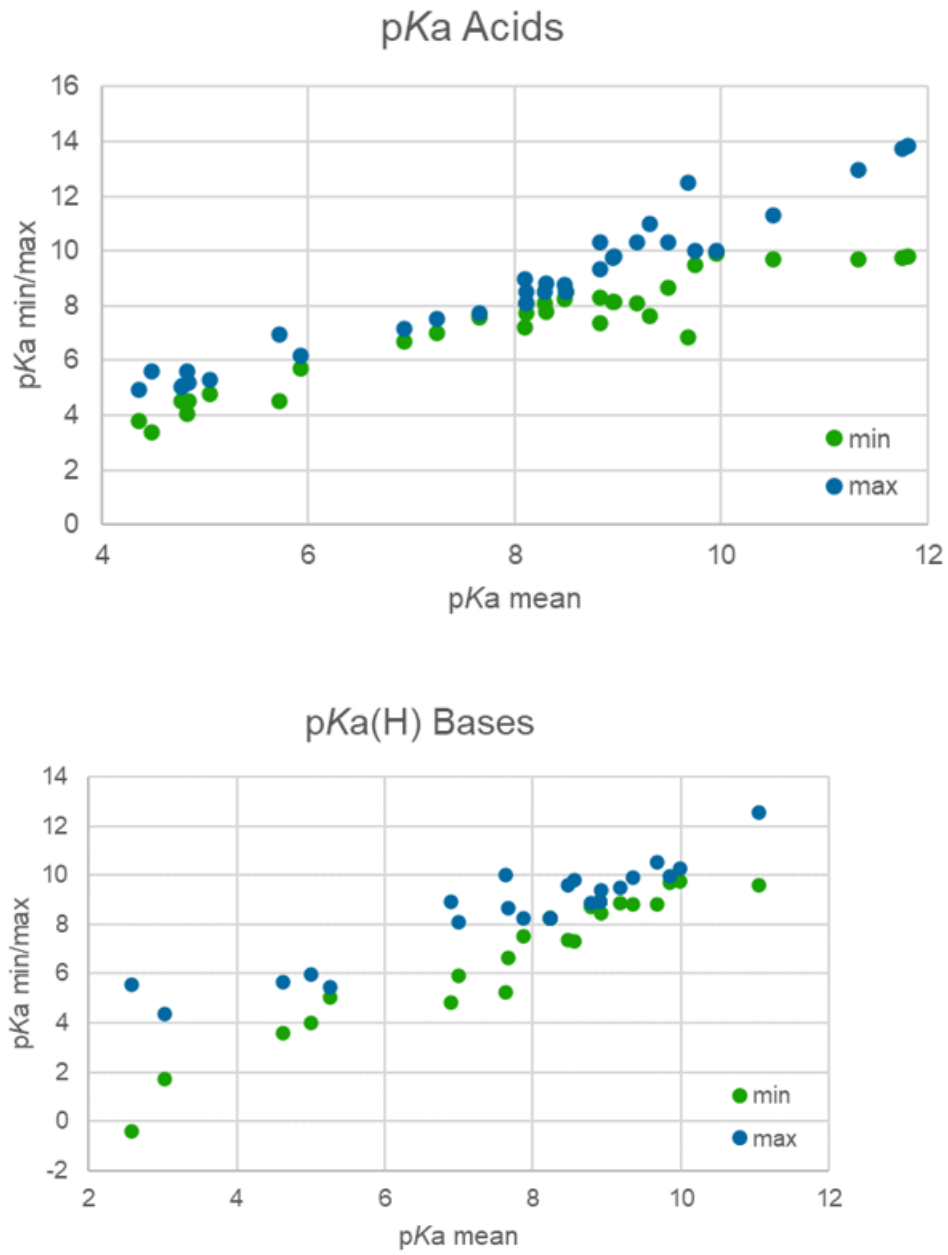
Case study (p*K*a): Results

The possible impacts of the dissociation of acids and bases on their hydrophobicity and water solubility, and thus on their fate and effects in the environment, can be assessed based on their p*K*a values. Compared to the numerous tools available for the prediction of log K_{OW} or log K_{OA}, access to computational methods for the calculation of p*K*a values is limited. Therefore, the data analyses and the resulting recommendations are also less comprehensive. In the following sections, first, several p*K*a values for the case study chemicals are presented, which were determined by different methods, and their data quality and variability are addressed. Then, the variability of the p*K*a estimates is discussed regarding the underlying methods and the different chemical classes. Finally, approaches are recommended to obtain reliable and robust p*K*a estimates for the hazard and risk assessment of chemicals.

Availability of experimental and calculated p*K*a values

In addition to experimental p*K*a values from different sources and of varying quality, calculated p*K*a values were available from two different conceptual methods, being either LFER (Hammet equation)-based (ACD/Labs 2023) or using automated read-across (k-nn) (Mansouri et al. 2018). For the 33 acids and 26 bases among the case study chemicals with p*K*a or p*K*a(H) between 4 and 10, between one and 18 data points could be obtained. The quality and evaluation of these data is presented and discussed below. Furthermore, observations and experiences with data handling are described.

Figure 51: Minimum and maximum p*K*a or p*K*a(H) values of the acids (upper panel) and the bases (lower panel), sorted by increasing mean values.



Source: own illustration, AL-Luhnstedt

Data quality of experimental and calculated p*K*a values

Major issues in data quality of experimental and calculated p*K*a values are related to incorrect documentation, e.g. the reporting of *K*a values instead of p*K*a values, or the mistaking of p*K*a (acidic) for p*K*a (basic), and vice versa. The former is found repeatedly in the ECHA database (ECHA 2023), while the latter is particularly seen with the ACD/Labs estimates in the CompTox Chemicals Dashboard (US EPA 2023).

In EAS-E Suite (ARC 2023), it is noticeable that substances without stored p*K*a values are incorrectly specified as 'neutral at pH 7.4'. Examples of this are isophthalic acid (CAS 121-91-5) and nonylamine (CAS 112-20-9). Overall, our experience from this case study is that the

identification of reliable p*K*a values requires expertise so that incorrect or misleading data can be excluded.

Variability of experimental and calculated p*K*a values

Appropriate curation of the available data for the 33 acids and 26 bases among the case study chemicals with p*K*a or p*K*a(H) between 4 and 10, results in average ranges of 1.5 log units. For about half of the acids (17 (52 %)) and bases (13 (50 %)), these p*K*a values are within 1 log unit, while for about one third of the acids (9 (27 %)) and bases (9 (35 %)) they vary by more than 2 log units. Figure 51 illustrates this variability of the p*K*a or p*K*a(H) values of the acids (upper panel) and the bases (lower panel). In both cases, some values exceed or fall below the range of 4 to 10, either the minimum values in the lower range and the maximum values in the upper range.

Based on the respective p*K*a values, ACD Galas (ACD/Labs 2023) also gives the percentage of the neutral species of an acid or base at pH 7 according to the Henderson-Hasselbalch relationship. In comparison, OPERA (Mansouri et al. 2018) only offers a classification of ionisation as 0 (neutral) or 1 (dissociated). The corresponding information from EAS-E Suite has already been assessed as not useful in the data quality section.

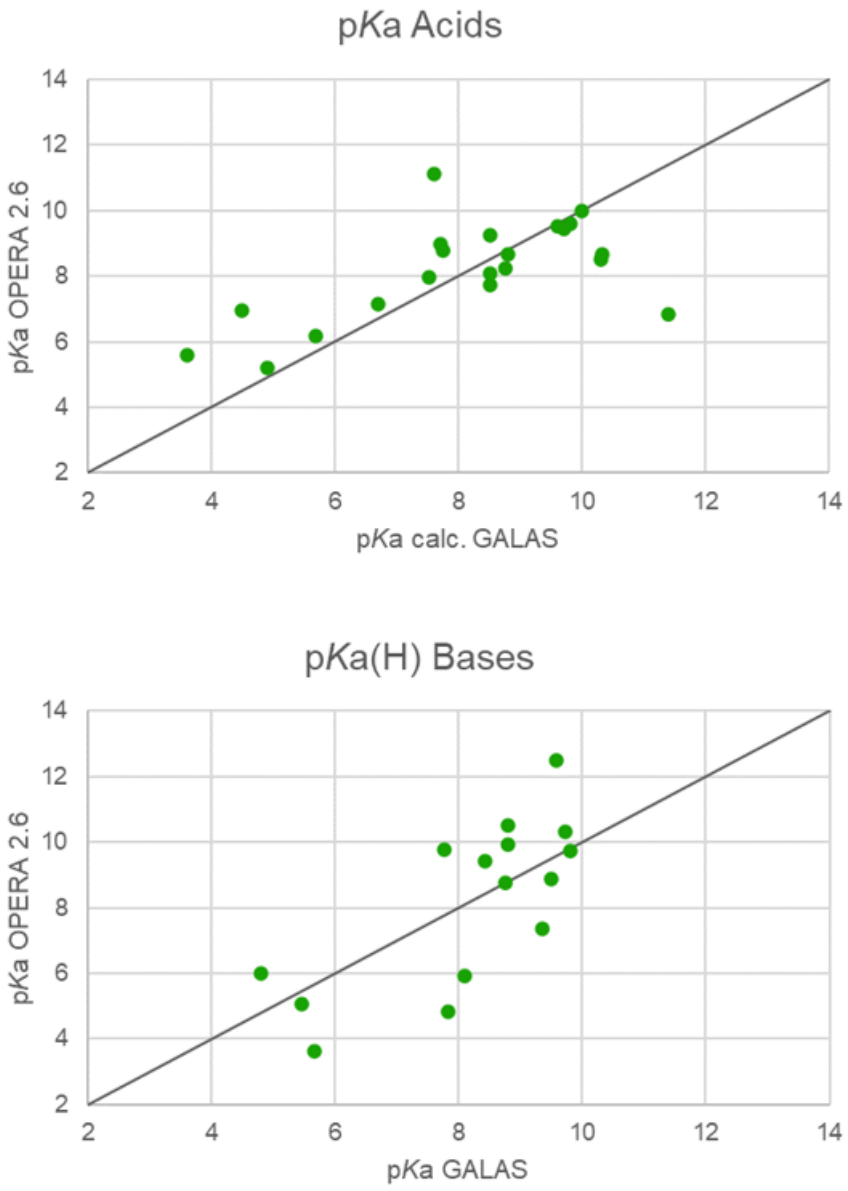
For weak acids and bases, the question arises as to the pH-dependent relevance of ionisation. For example, the dissociation of weak acids, e.g. with p*K*a 9, in alkaline waters with pH 8 is no more than 10% and thus has only a minor influence on their bioaccumulation behaviour. This influence further decreases with increasing p*K*a of the acidic substance or pH of the water body. For alkaline substances, the same dependencies between the p*K*a(H) of the substance and the ambient pH in the water body hold, however, in the opposite direction.

Accordingly, the question of the significance of the variability of p*K*a or p*K*a(H) values for the assessment of the bioaccumulation behaviour of acids and bases depends on the scenario under consideration.

Evaluation of the variability of p*K*a estimates with regard to the underlying methods and in terms of different chemical classes

To analyse the performance of the different methodological approaches for estimating p*K*a or p*K*a(H) between 4 and 10, the calculated values obtained with LFER (Hammet equation)-based methods using ACD/Labs GALAS (ACD/Labs 2023) and automated read-across (k-nn) using OPERA 2.6 from the CompTox Chemicals Dashboard (US EPA 2023) were compared (Figure 52). The agreement between the estimates is reasonable for the acids with 4 major outliers (mono-n-butyl phthalate (CAS 131-70-4), pentachlorophenol (CAS 87-86-5), hexaflumuron (CAS 86479-06-3), lotilaner (CAS 1369852-71-0)). In contrast, the estimates for the bases scatter without systematic pattern.

Figure 52: **Comparison of p*K*a estimates for acids (upper panel) and p*K*a(H) estimates for bases (lower panel) by the LFER (Hammet equation)-based method using ACD/Labs GALAS (ACD/Labs 2023) and automated read-across (k-nn) using OPERA 2.6 from the CompTox Chemicals Dashboard (US EPA 2023). The solid line is the 1:1 line.**



Source: own illustration, AL-Luhnstedt

For well-defined chemical classes such as phenols or alkylamines among the case study chemicals, consistent p*K*a estimates are obtained (Figure 53). For other substance classes, however, such analyses are not possible due to the limited data set.

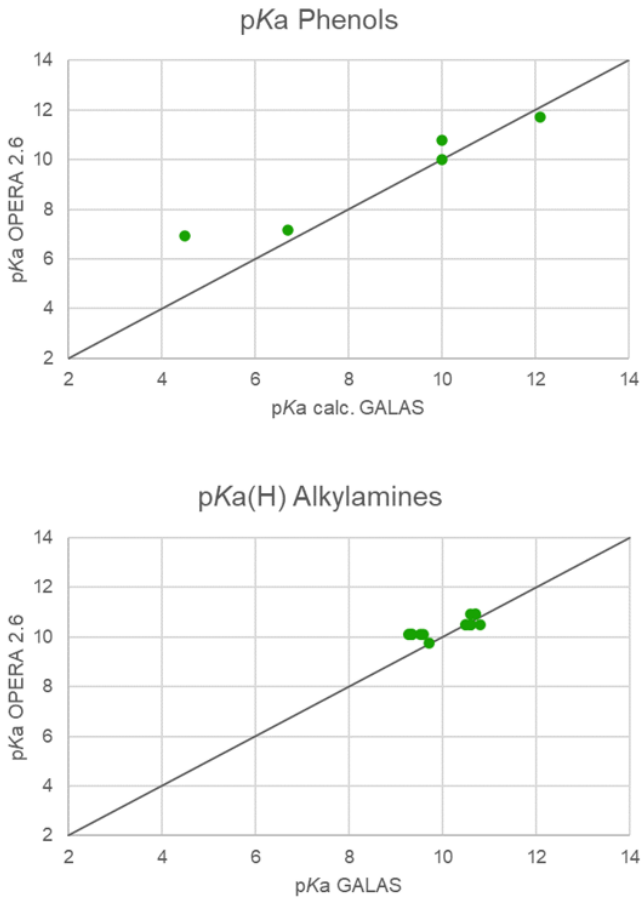
Zwitterionic and polyprotic substances, e.g. vancomycin and rifampicin, are not covered suitably by the available methods for estimating p*K*a or p*K*a(H) and require individual assessment on a case-by-case basis.

Recommendations for obtaining reliable and robust p*K*a estimates for the hazard and risk assessment of chemicals

If reliable measured p*K*a values are not available, consistent calculated data by two or more independent methods, e.g. by the LFER (Hammet equation)-based method using ACD/Labs GALAS (ACD/Labs 2023) and automated read-across (k-nn) using OPERA 2.6 from the CompTox Chemicals Dashboard (US EPA 2023), are the recommended alternative. Provided that the computational methods for p*K*a estimation fulfil the OECD criteria for scientific validity (OECD 2007)[3] and that the target compounds fit within their AD, the OECD QSAR Assessment Framework can be used to document reliable QSAR predictions for regulatory applications (OECD 2023b).

If no acceptable p*K*a values can be obtained in this way, there is still the option of a worst-case approach to the bioaccumulation assessment of the non-dissociated substance. Further data requirements may arise only if thresholds are exceeded.

Figure 53: Comparison of p*K*a estimates for phenols (upper panel) and p*K*a(H) estimates for alkylamines (lower panel) by the LFER (Hammet equation)-based method using ACD/Labs GALAS (ACD/Labs 2023) and automated read-across (k-nn) using OPERA 2.6 from the CompTox Chemicals Dashboard (US EPA 2023). The solid line is the 1:1 line.



Source: own illustration, AL-Luhnstedt

[3] OECD criteria for the scientific validity of QSAR models: (i) a defined endpoint; (ii) an unambiguous algorithm; (iii) a defined domain of applicability; (iv) appropriate measures of goodness-of-fit, robustness and predictivity; and (v) a mechanistic interpretation, if possible (OECD 2007).

Acknowledgments

Thanks! to Thomas Herbst (AL, Luhnstedt, Germany), Nadin Ulrich (UFZ, Leipzig, Germany) and Karsten Nödler (TZW, Karlsruhe, Germany) for contributing p*K*a estimates.

A.13 Workshop Summary

The workshop *"New Strategies for Regulatory Bioaccumulation Assessment"*, held on March 11–12, 2025, Berlin, Germany, brought together experts from regulatory bodies, research institutions, and industry to address the pressing need for modernised approaches in evaluating the bioaccumulation potential of chemicals. In the face of evolving scientific knowledge, technological progress, and political goals such as the reduction of animal testing, the event aimed to explore how innovative methods and mechanistic understanding could be integrated into existing regulatory frameworks. The overarching goal was to ensure that bioaccumulation assessments remain scientifically robust, ethically sound, and fit for purpose in light of emerging challenges.

A central focus was placed on New Approach Methodologies (NAMs), including in vitro assays like OECD Test Guideline 319 and the HYBIT invertebrate-based test system. While these methods offer significant potential - particularly in reducing reliance on vertebrate testing - their regulatory implementation still remains limited. Workshop participants emphasised that NAMs must be further validated, their applicability domains precisely defined, and their results consistently interpretable in relation to traditional in vivo endpoints such as the fish BCF (OECD TG 305). It became clear that without clear frameworks and defined applications, NAMs are not yet ready to fully replace traditional methods in regulatory practice.

Another major theme of the workshop concerned the challenge of data variability and uncertainty in predictive models. Participants stressed the importance of providing structured guidance on model selection to avoid arbitrary use or cherry-picking of results. Some suggested a consensus modelling approach that should be based on a limited number of well-justified QSAR models, ideally three to five, and each model's applicability domain must be clearly understood and respected. Others maintained that this approach is too complex and should not be seen as a standard procedure. The development of harmonised criteria for interpreting divergent model outputs was seen as a critical need for improving scientific rigor in regulatory contexts.

The workshop also turned attention to complex and difficult-to-assess substance groups, particularly superhydrophobic chemicals, PFAS, and surfactants. These substances challenge conventional assessment paradigms: For superhydrophobics, dietary studies and kinetic modelling approaches may offer the most viable data if bioconcentration testing or in vitro testing are no longer validly feasible. In the case of PFAS and surfactants, mechanisms like protein binding and membrane interactions must be accounted for. A consistent conclusion was that current models and test guidelines often fall short for these substance classes, necessitating refined definitions and tailored strategies.

Bioaccumulation in air-breathing organisms emerged as a distinct discussion area. Here, in vitro-derived half-lives were generally seen as more robust than biomagnification factors (BMFs), due to reduced variability. However, for use in complex organisms, these values must be supported by physiologically based toxicokinetic (PBTK) modelling and validation against IVIVE systems. Participants also encouraged leveraging existing human data to inform and contextualise regulatory assessments and to develop and validate a specific test guideline.

Throughout the workshop, the need for regulatory harmonisation was repeatedly emphasised. Differences in trigger values, endpoints, and interpretation criteria across frameworks such as REACH, CLP, biocides, plant protection products and pharmaceuticals were seen as obstacles to coherent and consistent decision-making. The goal must be to develop science-based, pragmatic, and transparent schemes that are applicable across sectors. In this context, the interactive "World Café" format of Day 2 contributed specific proposals for revising tiered testing strategies and refining decision points. It was pointed out that the focus of the developed approaches is on hazard-based bioaccumulation assessment (decision B, vB or not B) but the risk-based assessment (e.g. secondary poisoning) has not been taken into account so far.

In conclusion, the workshop was recognised as a timely and necessary step toward building a more integrated and future-oriented system for chemical bioaccumulation assessment. While several open questions remain - particularly regarding difficult substances and model applicability - participants agreed on key priorities: improving method validation, clarifying testing strategies, and fostering cross-sectoral regulatory alignment. The insights and discussions generated during the event are expected to feed into ongoing initiatives at OECD, ECHA, and national authorities, supporting a science-driven transition toward more sustainable and animal-free chemical regulation.