scientific **data**

OPEN

DATA DESCRIPTOR

Accumulation-depuration data collection in support of toxicokinetic modelling

Aude Ratier & Sandrine Charles ✉

Regulatory bodies require bioaccumulation evaluation of chemicals within organisms to better assess toxic risks. Toxicokinetic (TK) data are particularly useful in relating the chemical exposure to the accumulation and depuration processes happening within organisms. TK models are used to predict internal concentrations when experimental data are lacking or difficult to access, such as within target tissues. The bioaccumulative property of chemicals is quantified by metrics calculated from TK model parameters after fitting to data collected via bioaccumulation tests. In bioaccumulation tests, internal concentrations of chemicals are measured within organisms at regular time points during accumulation and depuration phases. The time course is captured by TK model parameters thus providing bioaccumulation metrics. But raw TK data remain difficult to access, most often provided within papers as plots. To increase availability of TK data, we developed an innovative database from data extracted in the scientific literature to support TK modelling. Freely available, our database can dynamically evolve thanks to any researcher interested in sharing data to be findable, accessible, interoperable and reusable.

Background & Summary

The Environmental Risk Assessment (ERA) workflow for chemical substances of interest, as described in some European regulations (*e.g.*, for plant protection products in marketing authorisation applications (EU regulation No 283/2013)), requires a bioaccumulation test, for example on fish according to the OECD Test guideline 305[1]. Such a test consists in an accumulation phase followed by a depuration one and the time course of the internal concentration within fish is measured at regular time points during both phases. The resulting data allow us to model the time-course of the exposure within organisms (denoted toxicokinetic, TK), summarized in the end via bioaccumulation metrics (BCF/BSAF/BMF for water, sediment and food exposure, respectively). From a regulatory point of view, these bioaccumulation metrics are key decision criteria to determine the bioaccumulative property of chemical substances[2], and to further assess potential risks that are associated with them according to the exposure sources.

Kinetic bioaccumulation metrics are always defined as ratios between uptake and elimination rates, these latter being estimated from TK models[1,3]. More precisely, TK models relate the exposure concentration to a given chemical substance to the internal concentration within organisms, considering various processes such as absorption, depuration, metabolism and excretion (ADME)[4]. These last decades, different types of TK models have been proposed, all being compartment models[5,6], where organs are considered as biological compartments connected through a fluid, usually blood. Organisms are thus described as whole or divided into organs, with input and output fluxes whose dynamics are described by TK models. Data collected from standard bioaccumulation tests are used to fit the TK models providing uptake and elimination rate estimates. They are also a way to provide a quantitative mechanistic framework to understand and simulate the time-course of the concentration of a chemical substance in various target organs accounting for body fluids in the case of physiologically-based toxicokinetic (PBTK) models. More generally, PBTK models allow to perform extrapolations that are inherent to risk assessment (*e.g.*, extrapolations from one species to another, between exposure routes, from one exposure scenario to other ones, …), and to calculate bioaccumulation metrics[5] thus helping decision-makers in the regulatory context.

Université de Lyon, Université Lyon 1, CNRS UMR5558, Laboratoire de Biométrie et Biologie Evolutive, 69100, Villeurbanne, France. ✉e-mail: sandrine.charles@univ-lyon1.fr
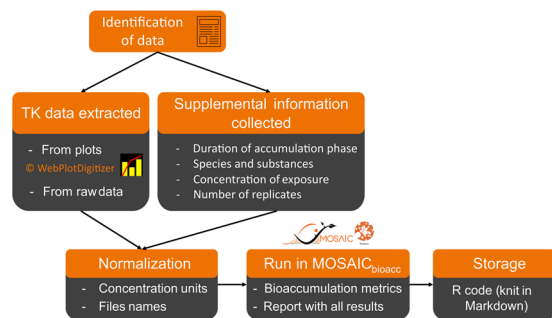
Fig. 1 Conceptual framework of the collection of TK data and their storage in the database of MOSAIC$_{bioacc}$.

Some years ago, the United States Environmental Protection Agency (US EPA) created a database for ecotoxicology data providing such bioaccumulation metrics according to species-compound combinations[7]. However, raw data are rarely provided in this database as this would require it to be collected directly from the corresponding scientific papers. Moreover, when available, raw data are mainly provided as plots. Nevertheless, to make updates or develop new TK modelling frameworks, it is crucial for researchers to be able to benefit from a collection of raw TK data in order to check the robustness of their new approaches. While developing the MOSAIC$_{bioacc}$ web application[8] (https://mosaic.univ-lyon1.fr/bioacc), we dealt with such a lack of raw data to fully test our innovative method. Indeed, MOSAIC$_{bioacc}$ provides estimates of TK model parameters and bioaccumulation metrics with their uncertainties for a large range of species-compound combinations (*e.g.*, aquatic or terrestrial organisms exposed to metals, hydrocarbons, active substances, etc.), encompassing different exposure routes and elimination processes. In particular, it was difficult to collect raw TK data with biotransformation processes being involved[9], preventing us from fully testing the robustness of MOSAIC$_{bioacc}$ for the widest possible diversity of input data types.

This motivated us in creating a new publicly available database as presented in this paper. This new database gathers together more than 200 datasets of published bioaccumulation tests, concerning more than 50 genus and more than 120 chemical substances. Some datasets concern several exposure routes (water, soil or sediment, and/or food) and several possible elimination processes (excretion, growth dilution and/or biotransformation). All the collected data were standardized in the same units and uploaded in MOSAIC$_{bioacc}$, ensuring the use of the same methodology in the acces of bioaccumulation metrics for all the datasets. This database should allow the current lack of raw TK data in ecotoxicology to be overcome. Indeed, our purpose with this first version of the database is to motivate other researchers to share their data based on the Findability, Accessibility, Interoperability and Reuse (FAIR) principles which have become today almost a duty[10]. Our database can be considered as a proof of concept of the added-value of sharing raw TK data. It allows anyone to reuse data, for example to test new modelling frameworks. This database should also facilitate the design of new bioaccumulation experiments, as well as the comparison of results between several studies, benefiting from the unified calculation method for bioaccumulation metrics as provided by MOSAIC$_{bioacc}$. Finally, we hope that this database could help in changing the paradigm in ecotoxicology, through an increase in a more wide sharing of raw data that would certainly lead to better knowledge in the perspective of ERA.

Methods

The conceptual workflow we followed to collect the raw TK data is summarized on Fig. 1 and detailed below.

Data and literature sources. One of our aims was to test the robustness of the MOSAIC$_{bioacc}$ methodological framework generic feature from a wide collection of TK datasets, encompassing a variety of genus, chemical substances, exposure routes and elimination processes. For this purpose, the US EPA ecotoxicology database[7] could not be used as raw observations are usually absent together with a lack of information (*e.g.*, biotransformation processes are usually not informed). Consequently, an in-depth search in literature with Scopus was performed with specific keywords according to the expected type of TK datasets. These keywords were for example: "TK model aquatic", "TK model terrestrial", "TK model biotransformation", "TK biotransformation rate", "TK model food exposure", "TK model sediment exposure", "TK model water exposure", etc (Table S1). For each retained article, we manually checked if TK data were available in tables or plots directly, in the manuscript or in the supporting information. A total of 56 studies were finally selected, fully analysed with a TK model and gathered in our database as a proof-of-concept. So, our corpus of data is therefore anything but exhaustive of what is available in the scientific literature.

Toxicokinetic data. *Literature source data extraction.* Only few publications provided raw data as directly usable tables of internal concentrations of a chemical (and its potential metabolite(s)) within organisms versus time. For most of the studies, data were only available in plots as dots to digitize. Plots were converted into JPEG files from screenshots (native screenshot tool in Windows 10). Then images were imported using WebPlotDigitizer[11] to extract the underlying numerical values corresponding to dots on plots. The calibration of axes and the selection of dots for extraction were performed manually. Finally, raw data are exported from WebPlotDigitizer to comma-separated values (CSV) files. Those CSV files of data were formatted following MOSAIC$_{bioacc}$ requirements, manually adding information on replicates and exposure concentration according

to the Methods section of each article. For each collected TK dataset, the corresponding file was named by the genus, the chemical substance, the duration of the accumulation phase, the first author and the year of publication, possibly adding comments (*e.g.*, number of tested organism) if available.

Data standardization. Measured data were standardized to always express exposure concentrations in $\mu g.ml^{-1}$ (for water exposures) or in $\mu g.g^{-1}$ (for sediment and food exposures). The internal chemical concentrations within organisms were translated into $\mu g.g^{-1}$, as required for a proper use of MOSAIC$_{bioacc}$.

TK modelling and inference framework under MOSAIC$_{bioacc}$. MOSAIC$_{bioacc}$ fits a one-compartment TK model written as a set of four ordinary differential equations (two per phases) to observed data. The inference process is implemented under a Bayesian framework based on a Monte-Carlo Markov Chain (MCMC, three chains by default) algorithm. The full model (and its generic solving, whatever the number of exposure routes and the number of phase I metabolites), as well as the inference method specifically adapted to get all parameter estimates simultaneously, are fully detailed in[12] and[13], respectively. New insights on the calculation of bioaccumulation metrics are provided in[8,14]. The user guide of MOSAIC$_{bioacc}$ that is available on-line from http://lbbe-shiny.univ-lyon1.fr/mosaic-bioacc/data/user_guide.pdf also explains the whole modelling process, together with details on the minimal data requirements that are expected to use MOSAIC$_{bioacc}$: at least one exposure concentration value, and time points with the corresponding measured internal concentrations; only accumulation data (during the accumulation phase of the test) are mandatory.

MOSAIC$_{bioacc}$ running. Once standardized, each collected dataset was uploaded into MOSAIC$_{bioacc}$ which automatically provided both kinetic and steady-state bioaccumulation metrics, unifying their calculation among datasets allowing relevant comparisons or a correct classification of chemical substances according to their bioaccumulative capacity for example. Kinetic bioaccumulation metrics (summarized as medians and 95% uncertainty intervals) were saved in the dedicated repository, and the report with all fitting outputs downloaded in the same repository, for each dataset. All calculations performed with MOSAIC$_{bioacc}$ can also be run directly in the R software[15] with the new rbioacc package[16].

Data Records

Storage and display. All collected datasets (directly downloadable as tabular files), the bibtex file with all references, all reports and all kinetic bioaccumulation metric estimates are publicly available on Zenodo[17]. An rmarkdown file[18,19] was created to build the overview table with information collected from the name of the dataset and from the dataset itself (*e.g.*, column headers, number of data, number of replicates), as well as from the bibtex file. The R package DT was additionally used[20] to combine all collected information in a user-friendly manner including a convenient search tool, and the rmarkdown file was finally compiled[19] in HTML format for display to the user in packs of 10 lines by default. In such a way, each new dataset added into the repository will compile the rmarkdown file automatically for update.

In parallel, the database can also be accessed directly via http://lbbe-shiny.univ-lyon1.fr/mosaic-bioacc/data/database/TK_database.html, or from MOSAIC$_{bioacc}$ clicking on the "More scientific TK data" button. An example of the output of the overview table is shown in Fig. 2, while the full table is provided in the supplementary information (Table S2). The collected raw TK data of the database consist in the time-course of several types of chemical substances bioaccumulated in various species via different exposure routes.

Datasets overview. Each dataset is summarized by:

- the file name (raw data directly downloadable by clicking on the file name, in text or CSV format),
- the genus of the tested organism,
- the category of the organism (*e.g.*, aquatic, terrestrial, etc.),
- the tested chemical substance,
- the duration of the accumulation phase,
- the tested exposure routes (*e.g.*, water, sediment, food, pore water),
- the total number of observations in the dataset (plus the number of replicate(s) in brackets),
- the kinetic bioaccumulation metric median value with its 95% uncertainty interval,
- the report which contains all the outputs from MOSAIC$_{bioacc}$ (in PDF format),
- the link to the reference or the source of the data,
- some additional comments (*e.g.*, lipid fraction, growth, biotransformation, if exposure was done for chemical mixtures or not, if total radioactivity was used or not, etc.).

A summary of all datasets is presented in Table 1. Genus were separated in 12 categories: aquatic invertebrates ($n = 105$), fish ($n = 42$), insects ($n = 17$), aquatic worms ($n = 10$), terrestrial worms ($n = 16$), seawater sponges ($n = 2$), seawater plants ($n = 1$), aquatic algae ($n = 1$), terrestrial invertebrates ($n = 1$), vertebrates other than fish ($n = 4$), marine invertebrates ($n = 8$), and heterotrichea ($n = 4$). The most represented genus in the database are *Gammarus* (aquatic invertebrate, $n = 43$) and *Daphnia* (aquatic invertebrate, $n = 27$), followed by *Oncorhynchus* (fish, $n = 15$), genus that are classically used in ecotoxicological experiments. Recommended genus by OECD guidelines for bioaccumulation tests are *Eisenia* and *Enchytraeus* for terrestrial organisms (OECD 317)[21], and *Tubifex* or *Lumbriculus* for aquatic invertebrates exposed to sediment (OECD 315)[22]; some datasets for these specific species are available in the database ($n = 24$).

Chemical substances were divided in 10 classes following at the best the nomenclature used in Standartox[23]: pesticides ($n = 71$), hydrocarbons ($n = 37$), metals ($n = 20$), nanoparticules ($n = 23$), polychlorobiphenyls (PCB,

Database

Show 10 ∨ entries   Search: [   ]

dataset	genus	category	substance	duration of accumulation	exposure route	n (replicate)	Kinetic biaccumulation factor	report	reference	comments
Anax_chlorpyrifos_2.313d_Rubach2010.txt	*Anax*	aquatic invertebrate	chlorpyrifos	2.313 days	water	49 (5)	113 [82 - 181]	Link	Rubach *et al.* (2010)	total radioactivity
Apostichopus_azithromycin_28d_Zhu2020.csv	*Apostichopus*	marine invertebrate	azithromycin	28 days	water	19 (1)	81 [74 - 88]	Link	Zhu *et al.* (2020)	
Apostichopus_clarithromycin_28d_Zhu2020.csv	*Apostichopus*	marine invertebrate	clarithromycin	28 days	water	19 (1)	92 [80 - 104]	Link	Zhu *et al.* (2020)	
Apostichopus_enrofloxacin_28d_Zhu2020.csv	*Apostichopus*	marine invertebrate	enrofloxacin	28 days	water	19 (1)	75 [69 - 82]	Link	Zhu *et al.* (2020)	
Apostichopus_ofloxacin_28d_Zhu2020.csv	*Apostichopus*	marine invertebrate	ofloxacin	28 days	water	19 (1)	39 [34 - 45]	Link	Zhu *et al.* (2020)	
Apostichopus_sulfadiazine_28d_Zhu2020.csv	*Apostichopus*	marine invertebrate	sulfadiazine	28 days	water	18 (1)	6 [5 - 7]	Link	Zhu *et al.* (2020)	
Apostichopus_trimethoprim_28d_Zhu2020.csv	*Apostichopus*	marine invertebrate	trimethoprim	28 days	water	18 (1)	51 [46 - 57]	Link	Zhu *et al.* (2020)	
Asellus_cypermethrine_24h_Dalhoff2020.txt	*Asellus*	aquatic invertebrate	cypermethrine	24 hours	water	16 (2)	514 [422 - 629]	Link	Dalhoff *et al.* (2020)	
Asellus_chlorpyrifos_2.048d_Rubach2010.txt	*Asellus*	aquatic invertebrate	chlorpyrifos	2.048 days	water	48 (5)	3273 [2570 - 4558]	Link	Rubach *et al.* (2010)	total radioactivity
Axinella_Ag_7.04d_Genta2012.txt	*Axinella*	seawater sponge	Ag	7.04 days	water	66 (3)	3267 [2558 - 4285]	Link	Genta-Jouve *et al.* (2012)	in mixture, use of radiotracers

Showing 1 to 10 of 211 entries   Previous 1 2 3 4 5 … 22 Next

Fig. 2 Screenshot of the first page of the overview table of the database available from MOSAIC$_{bioacc}$.

Genus (n = 52)	chemical substances (n = 124)	Exposure route	Elimination process	Kinetic bioaccumulation metric (median)	Number of studies
freshwater invertebrate (n = 105)	metals (n = 20)	water (n = 137)	biotransformation (n = 34)	BCF < 1000 (n = 88)	56
fish (n = 42)	PCB (n = 22)	food (n = 38)	excretion (n = 211)	1000 < BCF < 2000 (n = 9)	
insect (n = 17)	pesticides (n = 71)	sediment (n = 36)		2000 < BCF < 5000 (n = 15)	
aquatic worm (n = 10)	flame retardants (n = 8)			BCF > 5000 (n = 25)	
terrestrial worm (n = 16)	pharmaceutical products (n = 14)			BSAF > 1 (n = 16)	
seawater sponge (n = 2)	hydrocarbons (n = 37)			BSAF < 1 (n = 20)	
seawater plant (n = 1)	octyphenols (n = 2)			BMF > 1 (n = 8)	
aquatic algae (n = 1)	PFAS (n = 7)			BMF < 1 (n = 30)	
terrestrial invertebrate (n = 1)	nanoparticles (n = 23)				
vertebrate (other than fish) (n = 4)	other (n = 7)				
marine invertebrate (n = 8)					
heterotrichea (n = 4)					

Table 1. Summary of the collected TK datasets. The classification of bioaccumulative properties are the same as in ECHA (2017)[2,25].

n = 22), flame retardants (brominated or chlorinated, n = 8), pharmaceutical products (n = 14), PFAS (n = 7), octyphenol (n = 2) and other (n = 7). Among all datasets, the majority of bioaccumulation tests were performed via spiked water (n = 137). Besides, 34 datasets account for biotransformation processes, considering from 1 to 8 metabolites.

According to ECHA (2017)[2], BCF below 1,000 means that the chemical substance is not bioaccumulative, whereas one ranging between 1,000 and 5,000 corresponds to a bioaccumulative chemical substance: low bioaccumulative if BCF ∈]1,000; 2,000]; mid-bioaccumulative if BCF ∈]2,000; 5,000]. If BCF is >5000, the chemical substance is classified as very bioaccumulative. These ranges are reported in Table 1, where BCF median estimates are >5000 for 25 datasets, indicating a very bioaccumulative capacity of the corresponding chemical substances for the corresponding genus. Concerning BSAF and BMF estimates, their value must be compared to threshold 1. A median BSAF estimate >1 indicates that the corresponding chemical substance can bioaccumulate from soil or sediment into organisms at the base of the non-aquatic food chain[24,25]; a median BMF estimate >1 indicates that the corresponding chemical substance can biomagnify in the trophic relationship under consideration[26]. In the database, 16 datasets in 36 led to BSAF >1, for genus *Eisenia* (n = 2), *Enchytraeus* (n = 6),
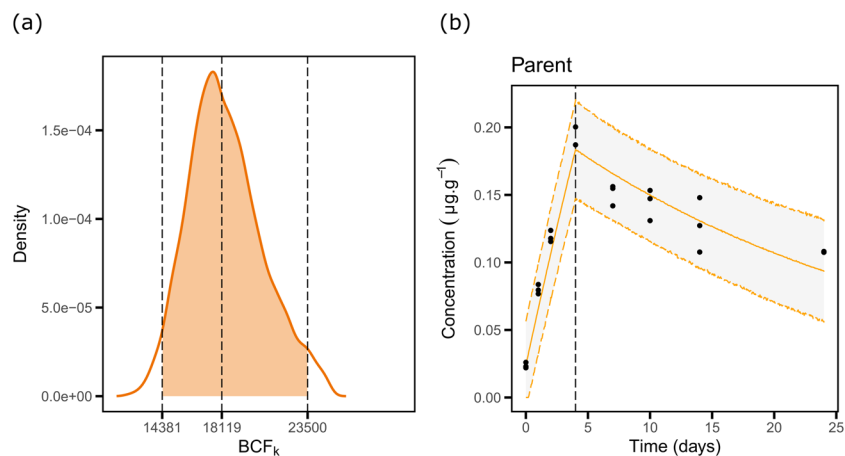
(a)

(b)

Fig. 3 Example of results given by MOSAIC$_{bioacc}$: (**a**) the bioaccumulation metric and (**b**) the fitted TK model predictions with observed data. This example comes from the dataset 'Male_Gammarus_Single.txt' available from MOSAIC$_{bioacc}$, where male gammarids were exposed to mercury (Hg) for 4 days in spiked water.

Gallus (n = 1), *Lumbriculus* (n = 2), *Metaphire* (n = 2), *Physa* (n = 1), *Radix* (n = 2)), while 8 datasets in 38 led to BMF >1, for genus *Gallus* (n = 1), *Oncorhynchus* (n = 5) and *Perca* (n = 2). On an ecotoxicological point of view, the highest BCF estimates were obtained for genus *Culex* and *Sialis* exposed to chlorpyrifos due to a very low estimate of the elimination rate, for genus *Gammarus* and *Calanus* exposed to hydrocarbons, and several aquatic invertebrates exposed to pesticides, especially chlorpyrifos (*n* = 4), attesting to the potential high bioaccumulation capacity and high risk of toxicity associated with this chemical substance for aquatic organisms. Overall, aquatic invertebrates seem to be the most sensitive category of organisms in terms of bioaccumulation of chemical substances representing 20 in the 25 datasets with a BCF estimates >5000.

Technical Validation

The validity of the collected data was checked by considering several aspects in the corresponding papers: (i) limitations in the material and method section, such as for example unreported numbers of replicates, avoiding to account for the variability between replicates; (ii) not enough details regarding the experimental design preventing reproducibility of the measurements (*e.g.*, only chemical name (source of ambiguity), the frequency of the renewal of the exposure media, etc.); (iii) the quality of the fitting results gotten from MOSAIC$_{bioacc}$ when reanalyzed, that is goodness-of-fit (GOF) criteria that were not satisfying enough, or low precision of parameter estimates.

For some studies, raw data were available in the supplementary files or upon request from the authors themselves, but appeared sometimes not consistent with the associated plots as provided within the original paper. For example, we sometimes detected some discrepancies in the plots, what led us to remove the corresponding datasets from our database. Besides, most of the studies did not provide the real raw data at all, such as separated concentration measurements in each replicate. Hence, most of the time, plots only provided average measurements with error bars standing for standard deviation, standard error or other uncertainty measures, thus preventing a full re-use of the data which are then reduced to mean values only.

The uncertainty due to data extraction from plots with the WebPlotDigitizer tool[11] is legitimate to account for as a bias may be introduced when digitizing dots. To quantify this bias, we performed the digitization of TK data from dots of a same plot (first saved as a .jpeg file) ten times and calculated the extraction errors possibly due to the shape of points, the axes calibration, the image resolution or other error sources. Hence, a variation of 0.12 ± 0.13% was quantified for the dataset of Chen *et al.*[27] (Fig. 2a in their paper), which can be considered as relatively low, given that values less than 0.5% are usually accepted for coefficients of variation. This result was representative of what we had for most of the plots, reinforcing our approach to digitizing the raw data.

In order to further promote the quality of the data in our database, MOSAIC$_{bioacc}$ was used to fit the appropriate TK models to all datasets, thus benefiting from the same conceptual framework and then make relevant comparisons. Indeed, bioaccumulation metrics and TK model fitting outputs were systematically provided (see an example on Fig. 3), as well as a collection of identical GOF criteria allowing for *a posteriori* check of the results. All this information was finally gathered together in a PDF report downloadable from the MOSAIC$_{bioacc}$ database for each dataset.

More than 95% of the datasets in the MOSAIC$_{bioacc}$ database were already published, meaning that they were publicly available for re-use. The re-analysis with MOSAIC$_{bioacc}$ allowed us to compare fitting results to those of the original study and to confirm the relevance of including the corresponding data in the database. In addition, very few of the publications (7 among 56) provided uncertainty around the predicted concentrations (at least, not visualized in the original plots), while we systematically associated the 95% credible intervals around median predicted estimates, as well as uncertainties for all bioaccumulation metrics, for each dataset. We are thus in full compliance with the recent recommendations from EFSA[28].

Usage Notes

Raw TK data are particularly useful to better understand the relationship between exposure chemical concentrations in the environment and their potential bioaccumulation within living organisms. They are also beneficial to researchers or regulators when evaluating the bioaccumulative capacity of chemical substances that is directly linked to potential toxic effects on life-history traits of the organisms. Our database also facilitates the access to tabular raw TK data that can be used to fit current TK models (with one or several compartment(s)), or to develop new conceptual modelling frameworks. It can also serve to simulate the time-course of the concentration of a chemical substance in various organs and bodily fluids in the case of PBTK models. Given that recommendations become more and more restrictive regarding the Replacement, Reduction and Refinement (3 R) principles[29], our database could reveal itself useful to perform extrapolations from one species, one exposure or one exposure scenario to any other ones of interest, in a change of paradigm in terms of model-guided experiments.

Our database does not represent an exhaustive overview of all available TK data in literature, but has the distinct advantage of encompassing several types of genus, chemical substances, exposure routes and elimination processes, thus constituting a supportive proof-of-concept of the feasibility and the major interest in sharing research data. Our database also aimed at developing and validating our generic TK modelling framework, as illustrated in this paper where the evaluation of the efficiency and robustness of MOSAIC$_{bioacc}$ proved to be successful for the whole set of collected data[9]. This is very much in line with Open Science as advocated today by most of research bodies[30].

When bioaccumulation testing is required by regulatory documents, it may be difficult to appropriately plan the experimental design which depends on the chemical substance, the species and the exposure route under consideration. Usually, a search in literature provides a possible range of exposure concentrations for the chemical substance, but the amount likely to be bioaccumulated in organisms of the chosen species remains unknown, as well as how long it will require to reach the steady-state at the end of the accumulation phase. In this perspective, the MOSAIC$_{bioacc}$ database quick search option (Fig. 2) may help to look at data on phylogenetically close genus and/or chemical substances with similar modes of action to drive new experiments.

Finally, our database could support further research to increase our knowledge on bioaccumulation of chemical substances for which there are clear societal challenges in terms of environmental protection today. In terms of regulatory ERA, bioaccumulation metrics are key decision criteria that classify a chemical substance as non-bioaccumulative, bioaccumulative or very bioaccumulative. Most often, these metrics are calculated from two or three different modelling approaches, which makes their comparison difficult. In addition, they are usually provided only as point mean values without uncertainty. Our database, together with the use of MOSAIC$_{bioacc}$, could contribute in improving this classification thanks to kinetic bioaccumulation metrics estimated with the same methodology for all datasets, with the quantification of the uncertainty summarized by the 95% credible intervals around median values. More globally, our database could encourage researchers to share their data with the scientific community by directly providing the raw data in their publications, as required by the FAIR data principles.

Code availability

In order to reproduce all the outputs provided in the PDF reports for all the datasets of the database: once downloaded, raw data can be uploaded in MOSAIC$_{bioacc}$ to perform the statistical analysis on-line. In addition, for each dataset, the corresponding R code can be retrieved from the download section of MOSAIC$_{bioacc}$, if using the R software is preferred.

References

1. OECD. Test no. 305: Bioaccumulation in fish: Aqueous and dietary exposure, https://doi.org/10.1787/9789264185296-en (2012).
2. ECHA. Guidance on Information Requirements and Chemical Safety Assessment. Chapter R.11: PBT/vPvB Assessment, https://echa.europa.eu/fr/guidance-documents/guidance-on-information-requirements-and-chemical-safety-assessment (2017).
3. Schlechtriem, C., Fliedner, A. & Schäfers, C. 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 (2012).
4. Mackay, D. & Fraser, A. Bioaccumulation of persistent organic chemicals: Mechanisms and models. *Environmental Pollution* **110**, 375–391, https://doi.org/10.1016/S0269-7491(00)00162-7 (2000).
5. Grech, A. *et al*. Toxicokinetic models and related tools in environmental risk assessment of chemicals. *Science of The Total Environment* **578**, 1–15, https://doi.org/10.1016/j.scitotenv.2016.10.146 (2017).
6. Landrum, P. F., Lydy, M. J. & Lee, H. II Toxicokinetics in aquatic systems: Model comparisons and use in hazard assessment. *Environmental Toxicology and Chemistry* **11**, 1709–1725, https://doi.org/10.1002/etc.5620111205 (1992).
7. USEPA. ECOTOX User Guide: ECOTOXicology Knowledgebase System. *Version 5.0*. https://cfpub.epa.gov/ecotox/help.cfm (2020).
8. Ratier, A. *et al*. New perspectives on the calculation of bioaccumulation metrics for active substances in living organisms. *Integrated Environmental Assessment and Management* **on-line**, https://doi.org/10.1002/ieam.4439 (2021).
9. Ratier, A., Lopes, C., Geffard, O. & Babut, M. The added value of bayesian inference for estimating biotransformation rates of organic contaminants in aquatic invertebrates. *Aquatic Toxicology* **234**, 105811, https://doi.org/10.1016/j.aquatox.2021.105811 (2021).
10. Wilkinson, M. D. *et al*. The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data* **3**, 1–9, https://doi.org/10.1038/2Fsdata.2016.18 (2016).
11. Rohatgi, A. Webplotdigitizer: Version 4.4 https://automeris.io/WebPlotDigitizer (2020).
12. Charles, S., Ratier, A. & Lopes, C. Generic solving of one-compartment toxicokinetic models. *Journal of Exploratory Research and Pharmacology* **on-line**, https://doi.org/10.14218/JERP.2021.00024 (2021).
13. Ratier, A. *et al*. A unified Bayesian framework for estimating model parameters for the bioaccumulation of organic chemicals by benthic invertebrates: proof of concept with PCB153 and two freshwater species. *Ecotoxicology and Environmental Safety* **180**, 33–42, https://doi.org/10.1016/j.ecoenv.2019.04.080 (2019).

14. Charles, S. *et al.* Taking full advantage of modelling to better assess environmental risk due to xenobiotics-the all-in-one facility mosaic. *Environmental Science and Pollution Research* https://doi.org/10.1007/s11356-021-15042-7 (2021).
15. R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria https://www.R-project.org/ (2021).
16. Ratier, A. *et al.* rbioacc: an R-package to analyse toxicokinetic data. *bioRxiv* https://doi.org/10.1101/2021.09.08.459421 (2021).
17. Charles, S. & Ratier, A. MOSAIC_bioacc database v2.0. *zenodo* https://doi.org/10.5281/zenodo.5032447 (2021).
18. Allaire, J., Horner, J., Xie, Y., Marti, V. & Porte, N. *markdown: Render Markdown with the C Library 'Sundown'. R package version 1.1* https://CRAN.R-project.org/package=markdown (2019).
19. Xie, Y. *knitr: A General-Purpose Package for Dynamic Report Generation in R. R package version 1.31* https://yihui.org/knitr/ (2021).
20. Xie, Y., Cheng, J. & Tan, X. *DT: A Wrapper of the JavaScript Library 'DataTables'. R package version 0.17* https://CRAN.R-project.org/package=DT (2021).
21. OECD. Test no. 317: Bioaccumulation in terrestrial oligochaetes https://doi.org/10.1787/9789264090934-en (2010).
22. OECD. Test no. 315: Bioaccumulation in sediment-dwelling benthic oligochaetes https://doi.org/10.1787/9789264067516-en (2008).
23. Scharmüller, A. *standartox: Ecotoxicological Information from the Standartox Database. R package version 0.0.1* (2021).
24. Weisbrod, A. V. *et al.* Evaluation of bioaccumulation using *in vivo* laboratory and field studies. *Integrated Environmental Assessment and Management* **5**, 598–623, https://doi.org/10.1897/IEAM_2009-004.1 (2009).
25. ECHA. Guidance on Information Requirements and Chemical Safety Assessment. Chapter R7.c: Endpoint specific guidance. Guidance for the implementation of REACH https://echa.europa.eu/fr/guidance-documents/guidance-on-information-requirements-and-chemical-safety-assessment (2017).
26. Gobas, F. A., de Wolf, W., Burkhard, L. P., Verbruggen, E. & Plotzke, K. Revisiting bioaccumulation criteria for pops and pbt assessments. *Integrated Environmental Assessment and Management* **5**, 624–637, https://doi.org/10.1897/IEAM_2008-089.1 (2009).
27. Chen, W., Li, W., Ju, Y., Liao, C. & Liao, V. Life cycle toxicity assessment of earthworms exposed to cadmium-contaminated soils. *Ecotoxicology* **26**, 360–369, https://doi.org/10.1007/s10646-017-1769-4 (2017).
28. Guidance on Uncertainty Analysis in Scientific Assessments. *EFSA Journal* **16**, 1–39 https://doi.org/10.2903/j.efsa.2018.5123 (2018).
29. Brandon, E. F. *et al.* Implementation of toxicokinetics in toxicity studies - Toxicokinetics of 4-methylanisole and its metabolites in juvenile and adult rats. *Regulatory Toxicology and Pharmacology* **73**, 55–64, https://doi.org/10.1016/j.yrtph.2015.06.018 (2015).
30. Dienlin, T. *et al.* An Agenda for Open Science in Communication. *Journal of Communication* **71**, 1–26, https://doi.org/10.1093/joc/jqz052 (2021).

Acknowledgements

Author contributions

A.R. and S.C. participated to the collection of datasets, analysed the results in MOSAIC$_{bioacc}$ and elaborated the database. All authors reviewed the manuscript and agreed on its content.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41597-022-01248-y.

Correspondence and requests for materials should be addressed to S.C.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.