R-BIONIC – An R package to Build Confidence in, and Drive Understanding of in Silico Predictions for Bioconcentration in Ionizable Compounds using Toxicokinetics

Patrik Engi [1], Tymoteusz Pietrenko [1], Jacob-Joe Collins [1], Dawei Tang [1], Ans Punt [1], and **Bruno Campos** [1]
[1] *Unilever Safety, Environmental & Regulatory Science, Colworth Science Park, Sharnbrook, Bedfordshire, MK44 1LQ, UK*
Email: patrik.engi@unilever.com

1. Background

There is regulatory and societal drive to move away from traditional toxicity testing to animal-free New Approach Methodologies (NAMs). An essential part of NAMs are in silico tools able to deliver robust predictions used to support safety decisions. The "BIONIC" model by Armitage et al. (2013) is a toxicokinetic model designed to predict bioconcentration factors (internal / external concentration ratios) for a wide range of chemical groups, including neutral, ionisable compounds (IOCs), and quaternary ammonium compounds (QACs).

2. Aims & Objectives

The overall aim of this work was to develop an R package that allows increased transparency and accessibility of the BIONIC v3 toxicokinetic model for predicting environmentally relevant bioconcentration factors for ionisable chemicals in fish. This can be broken down into two key objectives:

1) Development of an R package that translates the "BIONIC v3" model into R code.
2) Validation of the application of the R code against the "BIONIC" excel tool and its application for ionisable chemicals compared to a simple toxicokinetic model by Mackay et al.(2014)

3. Methods

Package Development

- Translation of the mathematical model described in Armitage et al. (2013) and excel VBA BIONIC tool into a R package, with a clear structure to the sub models for the model.
- We have included unit tests, and integration testing. With the latter being a set of benchmarking scenarios that test results against the model outputs from the VBA tool.
- Comprehension of the mathematical model to give clear documentation to each function
- Version controlled package that is user friendly and distributable in GitHub

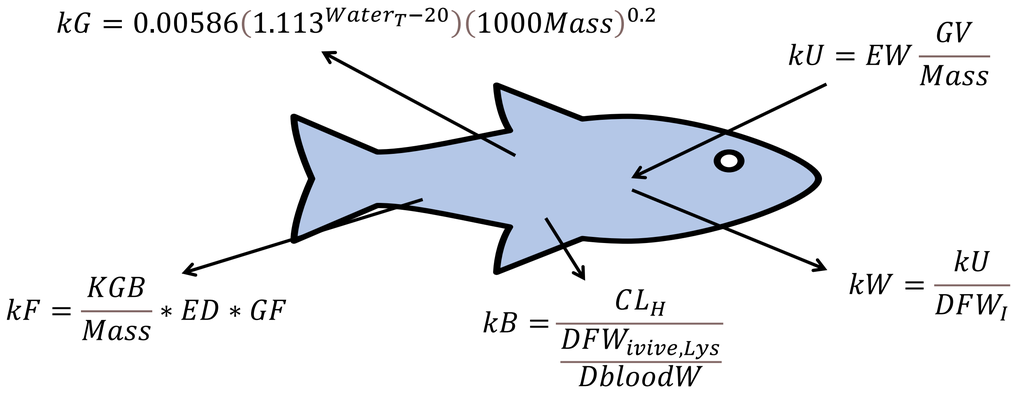
Mathematical Model

A one compartmental model made up of rate constants representing the different processes of elimination or uptake for a fish at steady state. The bioconcentration factor is given as:

$$BCF = \frac{C_B}{C_W} = V_w + \frac{kU}{kW + kB + kF + kG}$$

An Overview of the sub models is given as:

$$kG = 0.00586(1.113^{Water_T - 20})(1000Mass)^{0.2}$$



$BCF(L/kg)$ - the bio concentration factor, $C_B(g/kg)$ - the concentration of the chemical in the organism, $C_W(g/L)$ - the total concentration of the chemical in the water, $V_W(unitless)$ - the volume fraction of water, $kU\ (L/kg/days)$ - the Gill Uptake Rate Constant, $kW(1/days)$ - the Gill Elimination rate, $kG(1/days)$ - the growth rate constant, $kB(1/days)$ - the biotransformation clearance rate, $kF(1/days)$ - the fecal egestion rate constant, $EW\ (unitless)$ - chemical uptake efficiency at the gill, $GV\ (L/days)$ - ventilation rate, $Mass\ (kg)$ - the mass of the fish, DFW_I – Fish-water distribution ratio, $Water_T(℃)$ - the water temperature, $CL_H(L/kg/days)$ - Hepatic blood clearance, $DFW_{ivive,Lys}(L/kg)$ - Fish-water distribution ratio (IVIVE), adjusted for lysosome sequestration, $DbloodW\ (L/L)$ - Blood-water distribution ratio, $KGB(unitless)$ - The ratio of the partitioning into the gut divided by the total partitioning to the fish, $ED(unitless)$ - The absorption efficiency of chemical from food, $GF(kg/days)$ - the fecal egestion rate. It is important to note that this is the highest-level view of each sub-model, almost all the above parameters are themselves a products of other key inputs e.g. logP, pKa

Package Validation

- The package was benchmarked to the excel VBA BIONIC model using the same exemplar inputs into the package and VBA tool. Both gave matching outputs.
- The results of the model and sub-models are compared to be the same to validate the package

Data

- Data were gathered from a curated database of several thousand BCF measurements across a variety of species and chemicals (Arnot et al. 2008)
- The test set was narrowed first by selecting only measurements taken for vertebrates, then by choosing only the highest scoring for experiment quality (e.g. long exposure to satisfy steady-state model assumption) as provided by the original dataset.
- We further refined the test set of chemicals, by processing all chemicals using ACD/Labs (version 2024.1.5) to generate predictions for logP, measured pKa (and ionisation form). The labels for acids and bases were chosen based on the experimental ionisation form as designated by the reference source of the measured pKa values. Only chemicals with these descriptors were kept as these are minimum required inputs for the BIONIC model.
- All other model parameters were determined either using the information within the original (Arnot et al. 2008) database (e.g. temperature, or lipid fraction), or by assigning defaults using expert judgement.
- While the BIONIC implementation includes the capability to predict biotransformation rate (k_B) using S9 assay data, this was disregarded (assumed $k_B = 0$) as for this test set no such data were available.
- The data were used to not only to compare the accuracy of the prediction of the BCF to the true BCF for the package, but also to compare between the predictions of a simpler one compartmental model: The Mackay Model (Mackay et al. 2014)

4. Results/ Discussion

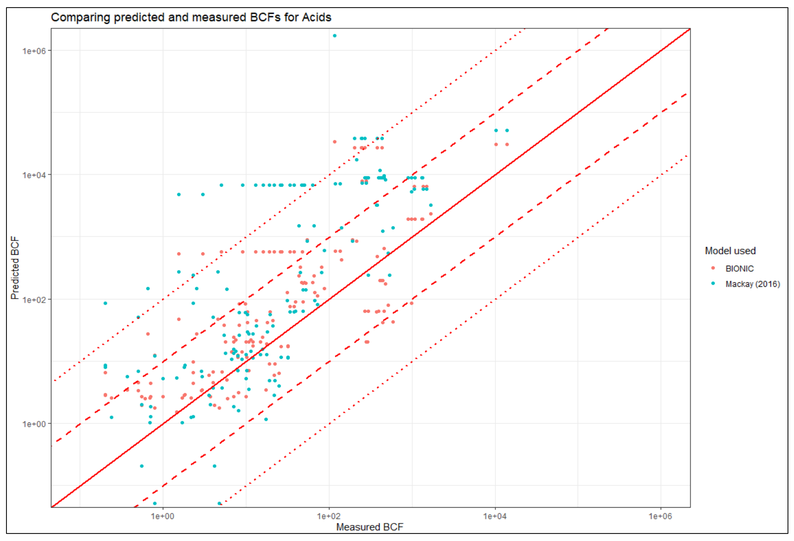


Figure 2: Comparison of log transformed predicted vs measured BCFs for a selection of organic acids using two different toxicokinetic models: BIONIC (as implemented in r-bionic) and Mackay's simplified NLOM model.

The results for the chosen test chemicals demonstrate that across both assumed acidic and basic chemical spaces there is an overall improvement in bioconcentration factor predictions when comparing the BIONIC model to the Mackay model. There is an approximate 0.4 log fold units or 2 times reduction to the average error for both groups. Where for acids BIONIC overpredicts to a lesser degree, and for bases the extent of underprediction is reduced.

The dotted lines show multiplicative (fold) differences of 10 and 100 when compared to the perfect prediction (solid line). We see that the majority of both models' predictions fall within a factor of 10 difference: with BIONIC being slightly better at 219/263 (83%) compared to 175/263 (67%) for Mackay. There are no BIONIC predictions which underpredict the BCF by a factor of 100, while Mackay has 13 predictions that exceed this.

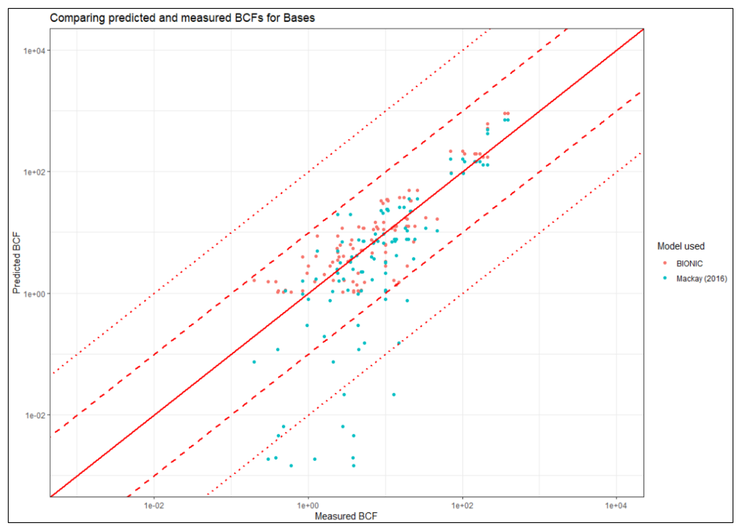


Figure 3: Comparison of log transformed predicted vs measured BCFs for a selection of organic bases using two different toxicokinetic models: BIONIC (as implemented in r-bionic) and Mackay's simplified NLOM model.

While the general goal of the model is to produce predictions closest to the measured values, for protective risk assessment purposes overpredictions are more beneficial, due to producing conservative predictions.

Statistic	BIONIC Acids	Mackay Acids	BIONIC Bases	Mackay Bases
Minimum	-1.1550	-3.2718	-1.0120	-3.4250
First quantile	0.0422	0.1554	-0.1859	-0.6409
Median	0.3512	0.7166	0.1396	-0.2041
Mean	0.4404	0.8448	0.0642	-0.4711
Third quantile	0.8325	1.5063	0.3520	0.1310
Maximum	2.5306	4.1630	0.9130	0.9158
Standard deviation	0.8222	1.1403	0.4056	0.9225

Table 1: Summary statistics for the distribution of log fold errors across the two models and ionisation types

By viewing some of the summary statistics in Table 1, we can determine that the distribution of errors for BIONIC show considerable improvement in model predictions. The standard deviation, range and interquartile range show that the spread of errors has been reduced, while mean and medians show them being closer to 0 when comparing against the Mackay model.

It is worth noting that biotransformation wasn't included within this round of analysis due to the lack of readily available data for this chemical space. We theorise that the BIONIC model predictions will further improve when including S9 assay data to predict biotransformation.

5. Conclusion / Future work

- Successfully translated the "BIONIC" v3 tool into an accessible and transparent R package for use by the toxicokinetic modelling community. Available once passed internal code review at https://github.com/seacunilever.
- Validation of R code outputs vs excel tool outputs highlighted accurate bioconcentration factors and key model parameters to 3 decimal places.
- Overall, the "BIONIC" model performed better in terms of a 2x reduction in average prediction error and more consistent predictions within 10-fold (83%) when compared with the Mackay model.
- Future work will focus on a sensitivity analysis that uses probabilistic methods to incorporate uncertainty to identify the most influential parameters driving model predictions. This will increase confidence in the model workings and enable us to understand the most important experimental data to collect.
- As biotransformation was not included because of data limitations it would be beneficial to validate the model by finding and incorporating new biotransformation data through *in vitro* or *in silico* predictions.
- Using the R package as a foundation future work will focus on improving and enhancing the model. This will include the ability to calculate reverse dosimetry for IVIVE applications.

References

1. Armitage, J.M., Arnot, J.A., Wania, F. and Mackay, D. (2013), Development and evaluation of a mechanistic bioconcentration model for ionogenic organic chemicals in fish. Environmental Toxicology and Chemistry, 32: 115-128.
2. Mackay, D., McCarty, L.S. and Arnot, J.A. (2014) Relationships between exposure and dose in aquatic toxicity tests for organic chemicals. Environmental Toxicology and Chemistry, 33: 2038–2046.
3. Arnot, J.A., and Gobas, F. A. (2004). A food web bioaccumulation model for organic chemicals in aquatic ecosystems. Environmental Toxicology and Chemistry, 23: 2343–2355.
4. Arnot J. A., Mackay D., Parkerton T.F., and Bonnell M. (2008), A database of fish biotransformation rates for organic chemicals. Environmental Toxicology and Chemistry, 27:2263– 2270

Unilever

seac.unilever.com