Predicting Bioconcentration Factor Using a Metabolism-Based Quantitative Structure-Activity Relationship Model

Julia Magliozzo, Nan An, and Adelina Voutchkova-Kostal
Chemistry Department, The George Washington University

Introduction

Bioconcentration refers to the absorption and retention of a substance through an animal's dermal and/or respiratory surfaces, excluding contributions from dietary uptake. Generally applied to aquatic organisms, it is quantified using a bioconcentration factor, defined as the ratio of substance concentration in the organism versus the water at steady state. In the case of fish, the bioconcentration factor can be considered the result of two competing processes: uptake of the substance through the gills and chemical elimination through respiratory exchange, fecal egestion, metabolic biotransformation of the parent compound, and growth dilution (Arnot and Gobas 2006). Aside from the direct implications for toxicity of xenobiotics in fish, the bioconcentration factor is also used as a midpoint in the determination of bioaccumulation factors needed to define water quality criteria for human health by EPA (EPA 2016) and as a criterion to designate substances as persistent, bioaccumulative and toxic (PBT) under the European Union's REACH regulations (Williams, Panko et al. 2009).

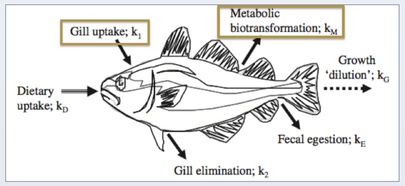


Figure 1. Major routes for chemical uptake and elimination in fish.

To meet these regulatory requirements, bioconcentration factor has traditionally been determined through *in vivo* testing. However, due to the time needed, cost, and number of fish used for testing (Gissi, Nicolotti et al. 2013) various *in silico* quantitative structure-activity relationship models (QSARs) have been developed as time- and cost-efficient alternatives. They include both linear and non-linear models applicable to a diverse range of substances.

Existing QSAR Models

- Linear based on octanol-water partition coefficient (logP or logKow)
- Non-linear based on logP
- VEGA CAESAR, Meylan, Read-across,
- EPISuite Meylan, Arnot-Gobas
- CORAL
- T.E.S.T.

Limitations of Existing Models

- High accuracy of prediction for compounds within their applicability domains but determining whether a compound falls within the applicability domain (AD) is not always straightforward.
- Significantly reduced performance when used to predict logBCF of compounds external to their training set (Gissi, Lombardo et al. 2015).
- Consider metabolism (or mechanistic interpretation) *a posteriori* to model development (Pavan, Netzeva et al. 2008). This is a significant oversight when the importance of mechanistic interpretation of QSAR models is highlighted in the Organization for Economic Co-operation and Development (OECD) guidelines for QSARs (Gissi, Gadaleta et al. 2014).

Metabolism-based Approach

To overcome the limitations of existing models Grisoni et al. (2015) suggested accounting for metabolism or processes other than lipid-driven accumulation patterns to develop models with high predictive accuracy and wide applicability domains.

The present work addresses this suggestion by using a subset of data (476 compounds total) from previous work by Dimitrov et al (2005) to develop a two-step linear QSAR model for bioconcentration factor. The two steps are:

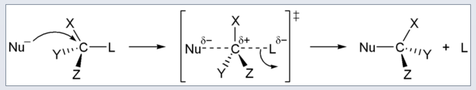
1. Tiered classification based on potential metabolism of the chemical. In this step chemicals are grouped based on structural characteristics as described by SMiles ARbitrary Target Specification (SMARTS) patterns.
2. Application of the appropriate QSAR given the metabolic classification.

The disadvantage of a linear model, the non-linearity of the logBCF-octanol-water partition coefficient relationship at very low and very high logP, is overcome by limiting the applicability domain to only those compounds with intermediate logD where linearity is observed (Wen, He et al. 2012). This approach is chosen to explicitly address mechanistic interpretations *a priori* to model development and shows valuable results.

Classification 1: Gill Permeation

Uptake through the gills is the primary mode by which chemicals are absorbed. At this interface the most important reactions should be those that alter electrophilic compounds as they pass through the gills before any metabolism occurs.

Scheme 1. A generic S_N2 reaction.



A common reaction that may occur is S_N2 substitution. In this reaction a good leaving group such as Cl, Br, or water is needed and is replaced by a nucleophile such as a hydroxide ion or an amine. This can alter the bioconcentration of the compound because the new group can change the interactions of the compound with the surrounding tissues.

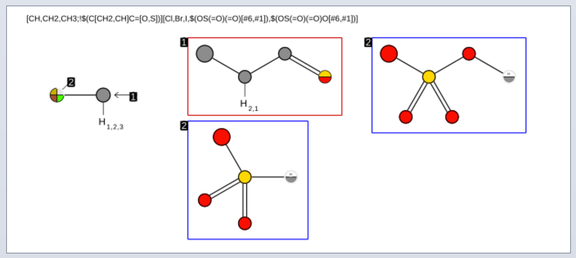


Figure 2. Example SMARTS for identification of potential gill permeation S_N2 reactivity. (Adapted from Kostal et al. (2016))

From the first classification, 72 compounds were identified as reactive. 404 had no match for initial gill permeation reactivity.

Classification, Continued

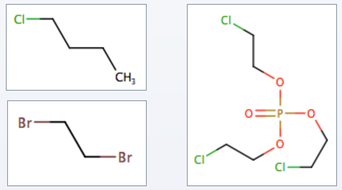
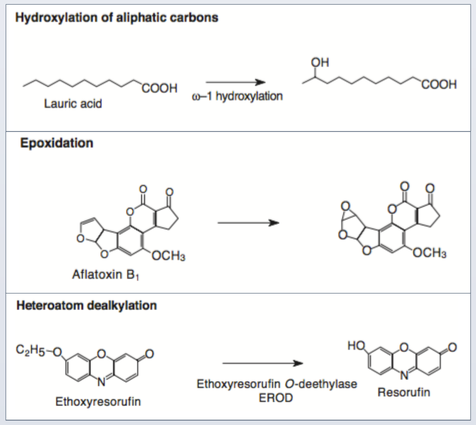


Figure 3. Examples of those compounds matched to the example SMARTS for S_N2 and classified as reactive at gill permeation.

Classification 2: Metabolism

The second mode of reaction of the chemicals is in metabolic pathways within the tissues of the fish. These reactions are catalyzed by enzymes, the most relevant of which we believe to be cytochrome P450. Cytochrome P450 is a very active and diverse enzyme that largely falls into the category of EC1 enzymes in the KEGG database. These enzymes are responsible for a breadth of reactions ranging from epoxidation to O,N-dealkylation to oxidation.

Scheme 2. Examples of known P450 catalyzed reactions. (Schlenk et al., 2008)



Compounds from our dataset will be classified as potential substrates for metabolism using SMARTS from previous work by Mu et al. Only those SMARTS pertinent to reactions catalyzed by EC1 enzymes will be used.

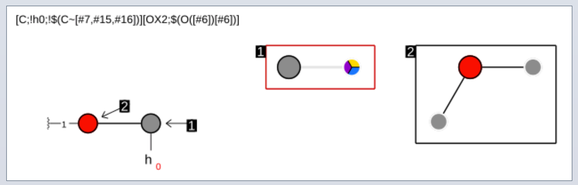


Figure 4. Example SMARTS used to classify chemicals based on potential metabolism. The SMARTS shown here flag compounds for O-dealkylation

QSAR Development and Next Steps

The dataset was limited to compounds whose logD values have a linear relationship with logBCF (logD = 0.5 to 6 for all chemical classes and LogD = 6 to 8 for halogenated biphenyls (Wen et al. 2012)) resulting in a final set of 403 compounds. After classifying the entire linear range dataset by gill permeation and metabolism we will establish whether there are meaningful differences between each class relative to the rest of the dataset and to each other. For example, preliminary data suggests that the gill permeation acyl substitution class shows a more constant logBCF given different values of logD.

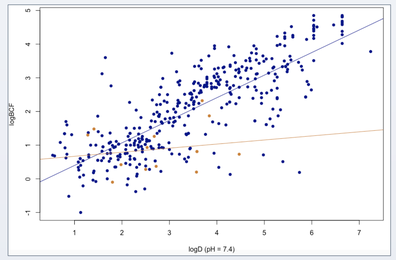


Figure 5. The entire linear range dataset is shown with compounds classified as ACYL shown in gold.

For each class (or group of classes) in which differences are found, a separate QSAR will be developed to capture this difference and potentially provide a more accurate prediction of logBCF.

Conclusions

We present here a systematic approach for including metabolism or mechanistic interpretation *a priori* to model development in a QSAR with a logBCF endpoint. We have not yet determined whether our two-step process for predicting logBCF will provide more reliable results than the existing QSAR models. However, regardless of the resulting QSARs, we believe the novelty of our two-tier classification scheme will fill a significant gap in the QSAR approach. Given the importance of considering metabolic processes in modeling bioconcentration we hope to provide a classification scheme that will allow for more knowledgeable and realistic bioconcentration QSARs to be developed in the future.

Literature Cited

1.Arnot, J. A. and F. A. P. C. Gobas (2006). "A review of bioconcentration factor (BCF) and bioaccumulation factor (BAF) assessments for organic chemicals in aquatic organisms." Environmental Reviews **14**(4): 257-297.
2.Dimitrov, S., et al. (2005). "Base-line model for identifying the bioaccumulation potential of chemicals." SAR and QSAR in Environmental Research **16**(6): 531-554.
3.EPA, U. (2016). Development of National Bioaccumulation Factors: Supplemental Information for EPA's 2015 Human Health Criteria Update. Office of Science and Technology. Office of Water. Washington, DC.
4.Gissi, A., et al. (2014). "An alternative QSAR-based approach for predicting the bioconcentration factor for regulatory purposes." Altex **31**(1): 23-36.
5.Gissi, A., et al. (2015). "Evaluation and comparison of benchmark QSAR models to predict a relevant REACH endpoint: The bioconcentration factor (BCF)." Environmental Research **137**: 398-409.
6.Gissi, A., et al. (2013). "Integration of QSAR models for bioconcentration suitable for REACH." Science of The Total Environment **456–457**: 325-332.
7.Grisoni, F., et al. (2015). "QSAR models for bioconcentration: Is the increase in the complexity justified by more accurate predictions?" Chemosphere **127**: 171-179.
8.Kostal, J.; Voutchkova-Kostal (2016). "A.M. CADRE-SS an in silico tool for predicting skin sensitization potential based on modeling of molecular interactions." Chem. Res. Toxicol **29**: 58-64
9.Pavan, M., et al. (2008). "Review of Literature-Based Quantitative Structure–Activity Relationship Models for Bioconcentration." QSAR & Combinatorial Science **27**(1): 21-31.
10.Petoumenou, M. I., et al. (2015). "Comparison between bioconcentration factor (BCF) data provided by industry to the European Chemicals Agency (ECHA) and data derived from QSAR models." Environmental Research **142**: 529-534.
11.Piir, G., et al. (2013). "Comparative analysis of local and consensus quantitative structure-activity relationship approaches for the prediction of bioconcentration factor." SAR and QSAR in Environmental Research **24**(3): 175-199.
12.Schlenk, D., et al. (2008). "Biotransformation in Fishes." In Di Giulio, R. T., Hinton, D. E., Eds.; The Toxicology of Fishes; CRC Press, Taylor and Francis Group, LLC: Boca Raton, FL; pp 153.
13.Wen, Y., et al. (2012). "Linear and non-linear relationships between bioconcentration and hydrophobicity: Theoretical consideration." Environmental Toxicology and Pharmacology **34**(2): 200-208.
14.Williams, E. S., et al. (2009). "The European Union's REACH regulation: a review of its history and requirements." Critical Reviews in Toxicology **39**(7): 553-575.

Acknowledgements

We would like to thank Dr. Bryan W. Brooks and Dr. Jone Corrales from CRASR at Baylor University for their insights in fish metabolism. We would also like to acknowledge Dr. Jakub Kostal for his assistance in the computational and SMARTS analysis.