Exploring the Features of the New EPI Suite™ Web Models and Applicability Domains Relevant to Environmental Toxicological Risk Assessment

Mary Kawa[a], Lauren Cassidy[a], Laura Morlacci[a], Wen-Hsiung Lee[b]
[a]SRC Inc.; [b]U.S. Environmental Protection Agency Office of Pollution Prevention and Toxics (OPPT)

ABSTRACT

Module specific enhancements were made to the U.S. Environmental Protection Agency's (EPA) EPI Suite™ graphical user interface (GUI) to facilitate evaluation of each model's domain of applicability and provide data for read-across or analogue selection. The incorporation of Analog Identification Methodology (AIM) fragments is especially important as it allows users to see the robustness of a given estimation model based on the available data sets. By displaying analogues in conjunction with fragment counts, the model user can bridge the uncertainty gap for data-poor chemicals. These updates were made under guidance by the EPA's Office of Pollution Prevention and Toxics (OPPT) which regulates both new and existing industrial chemicals under the Toxic Substances Control Act (TSCA), as amended by the Frank R. Lautenberg Chemical Safety for the 21st Century Act. Given the large number of chemicals submitted each year for PMN evaluation, EPI Suite™ is one of the key programs used to assist in developing assessments of the hazards and risks of these materials under time constraints. Now, with the newly deployed web version of this software package under development, scientists can convey the logic and rationale of the modeled results and applicability domain considerations with a simple approach.

APPLICABILITY DOMAINS

When estimating chemical properties using a quantitative structure-activity relationship (QSAR) model, the target chemical's suitability with the model must be evaluated. To aid in this evaluation, applicability domains (AD) are reported for many of the models in EPI Suite™. The AD of a QSAR model is the "response and chemical structure space in which the model makes predictions with a given reliability"[1]. Property estimations for target chemicals that are outside of the defined AD for a model are normally considered unreliable and are likely not suitable for regulatory decision-making[2]. In practice, ADs are defined using a combination of several criteria.

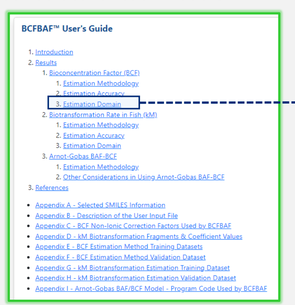
[1]OECD, 2007. Guidance document on the validation of (quantitative)structure-activity relationship [(Q)SAR] models.; Organisation for Economic Cooperation and Development. Paris.

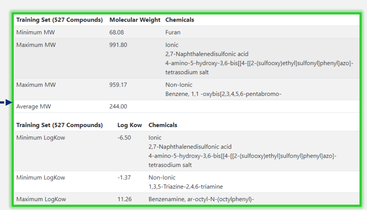
[2]Papa, E.; Sangion, A.; Arnot, J. 2018. Approaches to enhance the evaluation of the EPI Suite model(s) BCFBAF applicability domain.

GLOBAL APPLICABILITY DOMAINS

Global applicability domains establish the upper and lower bounds of a model's reliability based on a range of descriptors or responses for training set chemicals. Descriptors for the BCFBAF module include molecular weight and K_{OW}. Estimations for chemicals with a molecular weight or K_{OW} inside the range established by the training set are considered more reliable. The response global applicability domain is established by the range of experimental values present in the training set. If an estimated value is inside the range of measured values in the training set, the estimation is considered more reliable. These global domains, as well as any other applicability domain, are described in the User Guides for each module.

User Guide contents for BCFBAF module





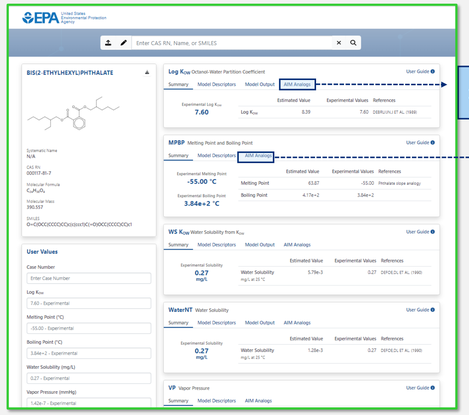
Training Set (527 Compounds)	Molecular Weight	Chemicals
Minimum MW	68.08	Furan
Maximum MW	991.80	Ionic 2,7-Naphthalenedisulfonic acid 4-amino-5-hydroxy-3,6-bis[[4-[[2-(sulfooxy)ethyl]sulfonyl]phenyl]azo]- tetrasodium salt
Maximum MW	959.17	Non-Ionic Benzene, 1,1 -oxybis[2,3,4,5,6-pentabromo-
Average MW	244.00	
Training Set (527 Compounds)	**Log Kow**	**Chemicals**
Minimum LogKow	-6.50	Ionic 2,7-Naphthalenedisulfonic acid 4-amino-5-hydroxy-3,6-bis[[4-[[2-(sulfooxy)ethyl]sulfonyl]phenyl]azo]- tetrasodium salt
Minimum LogKow	-1.37	Non-Ionic 1,3,5-Triazine-2,4,6-triamine
Maximum LogKow	11.26	Benzenamine, ar-octyl-N-(octylphenyl)-

Summary of global descriptor domains for BCFWIN

ANALOG IDENTIFICATION

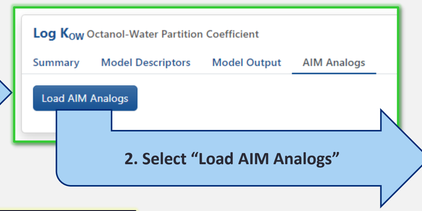
- AIM Analogs tabs have been added to the following EPI Suite™ modules, enabling users to quickly search the following training sets for analogs of the target chemical:
  - KOWWIN™
  - MPBPVP™
  - WSKOWWIN™
  - WATERNT™
  - HENRYWIN™
  - KOAWIN™
  - BIOHCWIN™
  - AEROWIN™
  - AOPWIN™
  - KOCWIN™
  - HYDROWIN™
  - BCFBAF™
  - WVOLWIN™
  - STPWIN™
  - LEV3EPI™
  - DERMWIN™
- The inclusion of multiple target chemical analogs in the training set indicates that the target chemical estimation is more reliable.
- EPI Suite™ results of the analogs can be run by clicking on the hyperlinked chemical name. This allows the user to quickly assess the estimation accuracy of the analog chemical against its experimental value, providing information about the model's estimation accuracy for chemicals in chemical space surrounding the target chemical.
- AIM is a software tool available at: https://www.epa.gov/tsca-screening-tools/analog-identification-methodology-aim-tool

EPI Suite™ Web Results Screen



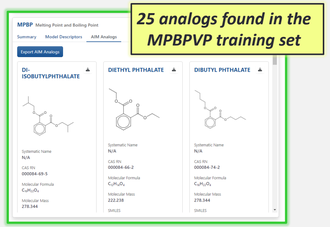
1. Open the "AIM Analogs" tab

Initial AIM Analogs Tab

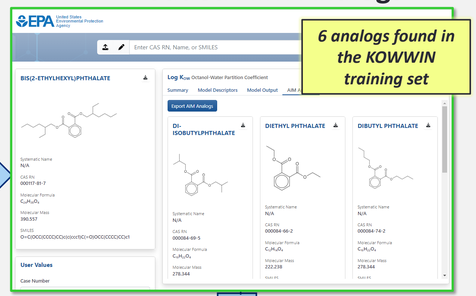


2. Select "Load AIM Analogs"

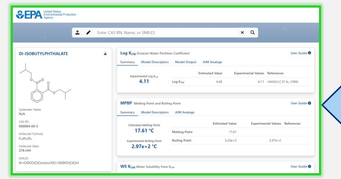
25 analogs found in the MPBPVP training set



Loaded AIM Analogs



6 analogs found in the KOWWIN training set



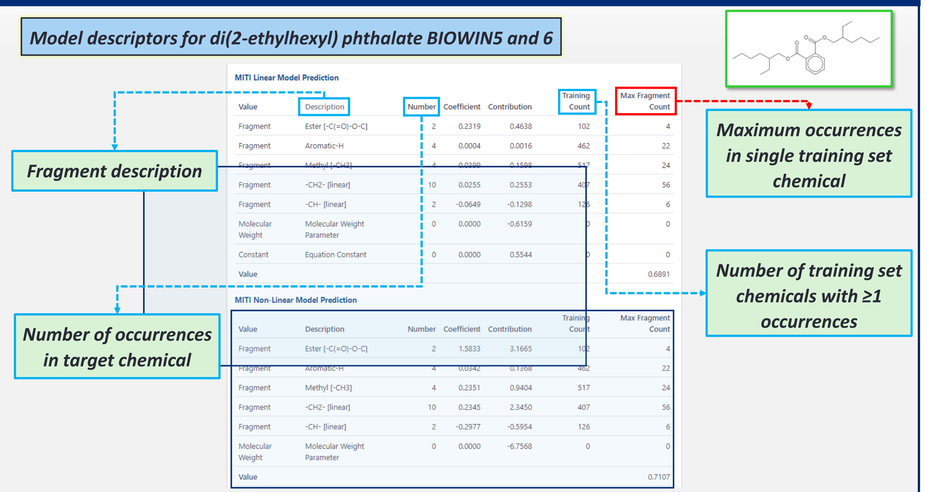
Hyperlinked chemical name leads to the full EPI Suite™ page of the analog chemical

FRAGMENT COUNTS

- The structural fragment domain is defined as the total set of fragments or functional groups present in the training set chemicals.
- If a target chemical contains a fragment or functional group that is not found in the training set, it is considered outside of the structural fragment domain and the estimation reliability is reduced.
- A target chemical may also fall outside of the structural fragment domain if it contains more instances of any fragment or functional group than is found in any training set chemicals (e.g., target chemical contains 8 -Br while the maximum occurrence found in a training set chemical is 7).
- When available, the Model Descriptors tab in the EPI Suite™ contains:
  - The number of chemicals containing each fragment of the target chemical (reported as "Training Count")
  - The maximum occurrence of each fragment in any training set chemical (reported as "Max Fragment Count")

  Users are directed to the model-specific User Guides for additional information regarding the structural fragment domains.

Model descriptors for di(2-ethylhexyl) phthalate BIOWIN5 and 6



MITI Linear Model Prediction

Value	Description	Number	Coefficient	Contribution	Training Count	Max Fragment Count
Fragment	Ester [-C(=O)-O-C]	2	0.2319	0.4638	102	4
Fragment	Aromatic-H	4	0.0004	0.0016	462	22
Fragment	Methyl [-CH3]	4	0.0299	0.1908	517	24
Fragment	-CH2- [linear]	10	0.0255	0.2553	407	56
Fragment	-CH- [linear]	2	-0.0649	-0.1298	126	6
Molecular Weight	Molecular Weight Parameter	0	0.0000	-0.6159	0	0
Constant	Equation Constant	0	0.0000	0.5544	0	0
Value						0.6891

MITI Non-Linear Model Prediction

Value	Description	Number	Coefficient	Contribution	Training Count	Max Fragment Count
Fragment	Ester [-C(=O)-O-C]	2	1.5833	3.1665	102	4
Fragment	Aromatic-H	4	0.0542	0.1968	462	22
Fragment	Methyl [-CH3]	4	0.2351	0.9404	517	24
Fragment	-CH2- [linear]	10	0.2345	2.3450	407	56
Fragment	-CH- [linear]	2	-0.2977	-0.5954	126	6
Molecular Weight	Molecular Weight Parameter	0	0.0000	-6.7568	0	0
Value						0.7107

Fragment description

Number of occurrences in target chemical

Maximum occurrences in single training set chemical

Number of training set chemicals with ≥1 occurrences

The views presented are those of the author(s) and do not necessarily represent the views or policies of the U.S. EPA