

Quantitative Structural activity relationships (Q)SAR's and their application to SDS and PIDS

John Morrison 23/11/16

(Quantitative) Structural activity relationships (Q)SAR's

1. Download suggested QSAR software (VEGA and ECHA/OECD QSAR toolbox) to see what they do.

2. What do we want to use them for? KEY QUESTION !!!

For Risk assessment of unknowns for transport no simple guidance found for use of the software packages to assist in our chemical transportation problems. (so several false starts throughout, particularly on ECHA).

3. Try worked examples:- Drop in our structures and see what comes out. Easy on VEGA but huge data mass. ECHA remarkably complex.

4. Assimilation of findings.

(Quantitative) Structural activity relationships (Q)SAR's

1. Download suggested QSAR software (VEGA and ECHA/OECD QSAR toolbox) to see what they do. Both free after giving a few details.

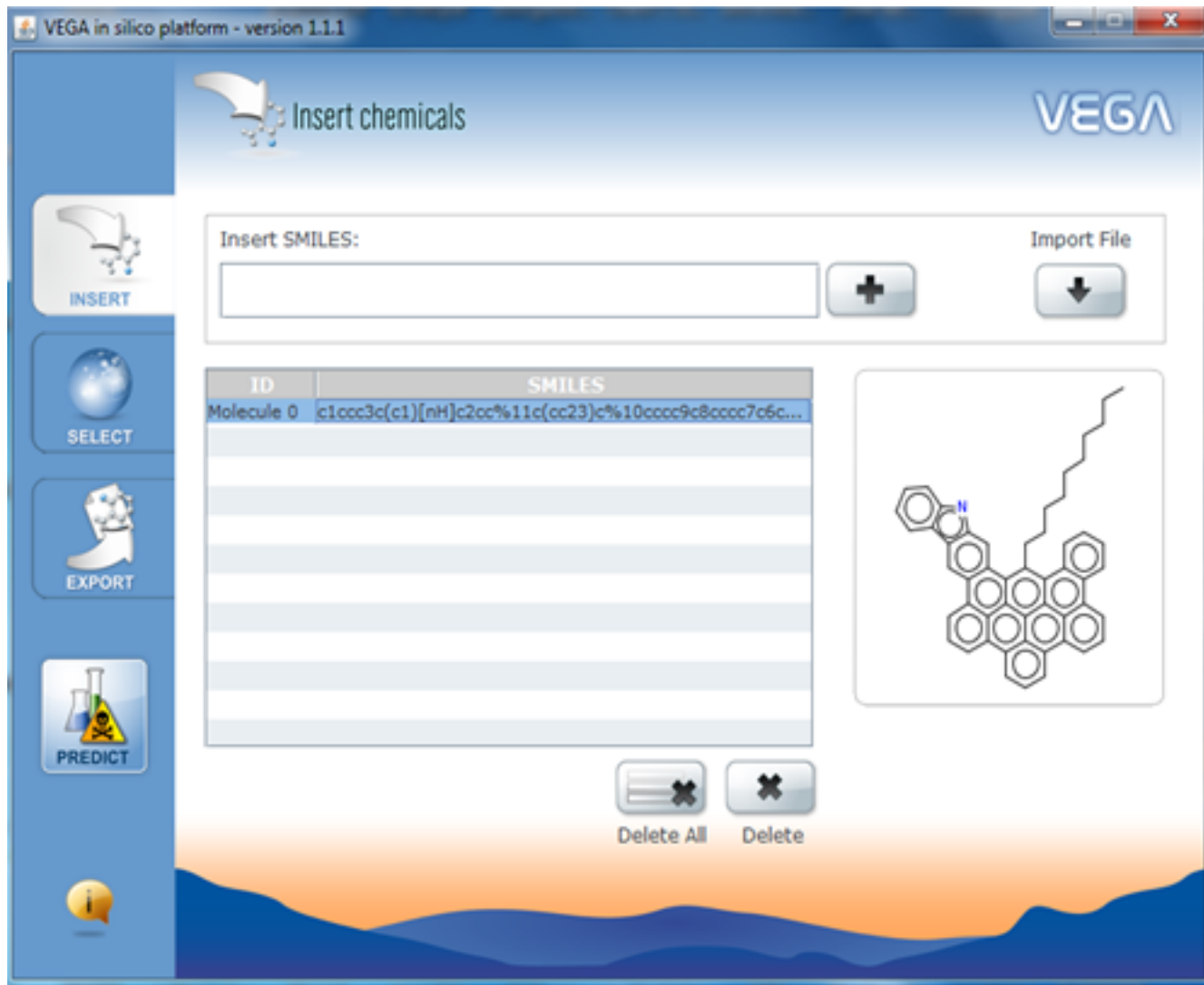
VEGA (<http://www.vega-qsar.eu/download.html>) is server based so not much to download. Got it working reasonably quickly then failed. Needs old version of Java to operate.

ECHA/OECD (<http://toolbox.oasis-lmc.org/?section=download&version=latest>) has server and standalone versions. Both are massive (3.7GB for server version). To date have not seen any tutorials for server version but many for stand alone.

For this would really need training eg

<http://oasis-lmc.org/products/software/toolbox/training.aspx>

VEGA (Q)SAR's

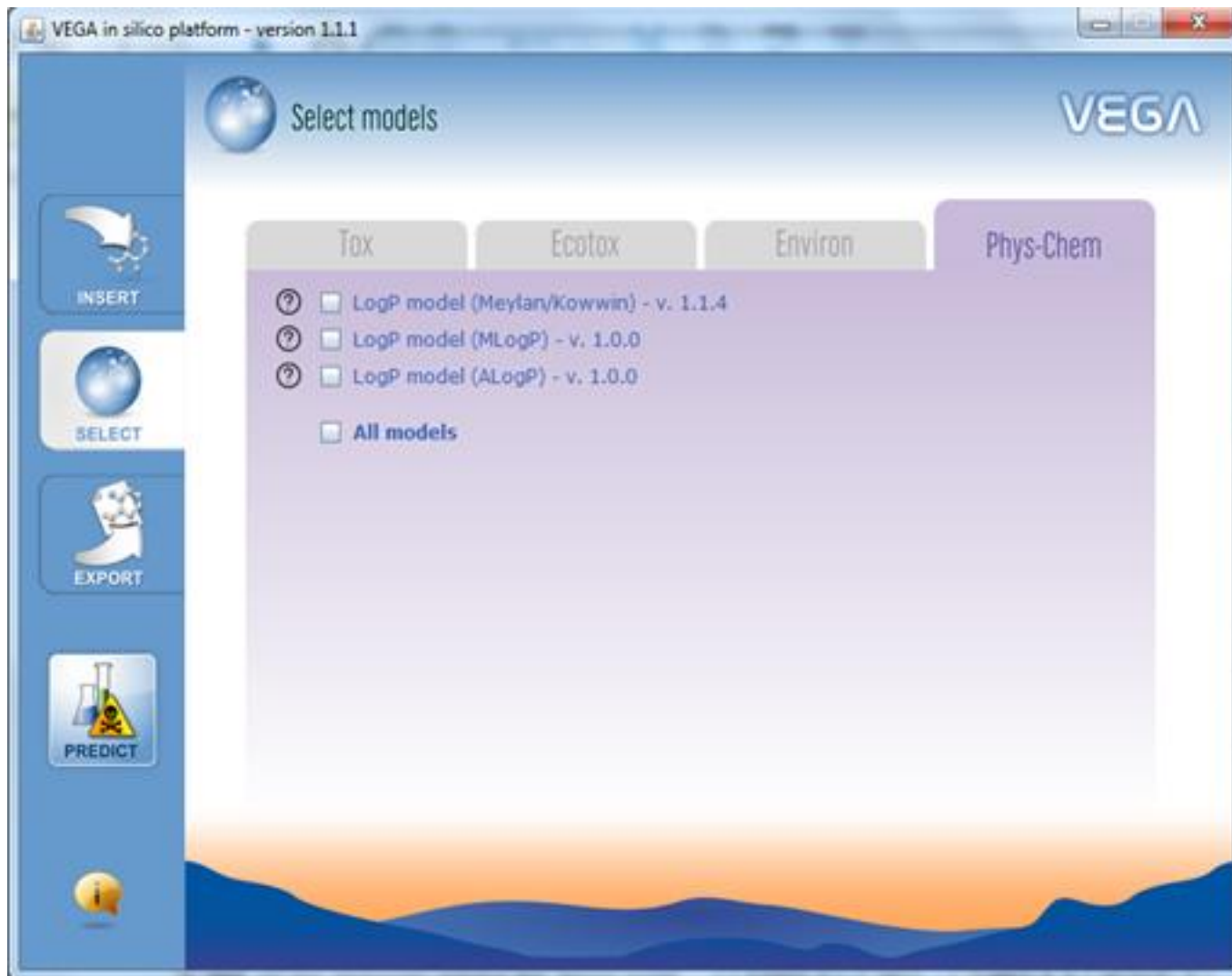


simplified molecular-input line-entry system (SMILES)

VEGA (Q)SAR's



VEGA (Q)SAR's



VEGA (Q)SAR's

VEGA

QSAR MODELS



Report



Prediction and Applicability Domain analysis for models:

Mutagenicity (Ames test) model (CAESAR) 2.1.13

Carcinogenicity model (CAESAR) 2.1.9

Developmental/Reproductive Toxicity library (PG) 1.0.0

Skin Sensitization model (CAESAR) 2.1.6

Fish Acute (LC50) Toxicity classification (SarPy/IRFMN) 1.0.2

Ready Biodegradability model (IRFMN) 1.0.9

LogP model (Meylan/Kowwin) 1.1.4

LogP model (MLogP) 1.0.0

LogP model (ALogP) 1.0.0

Core version: 1.2.1

9 Models gives 43 page report

VEGA (Q)SAR's

VEGA

Carcinogenicity model (CAESAR) 2.1.9

page 5



1. Prediction Summary

Prediction for compound Molecule 0

	Prediction: Reliability: Prediction is NON-Carcinogen, but the result may be not reliable. A check of the information given in the following section should be done, paying particular attention to the following issues: - only moderately similar compounds with known experimental value in the training set have been found - some similar molecules found in the training set have experimental values that disagree with the predicted value - predicted substance falls into a neuron that is populated by no compounds of the training set
--	---

Compound: Molecule 0

Compound SMILES:

c1ccc3c(c1)[nH]c2cc%11c(cc23)c%10cccc9c8cccc7c6cccc5c4cccc4c%12c(c56)c(c78)c(c9%10)c%11c%12CCCCCCCC

Experimental value: -

Predicted Carcinogen activity: NON-Carcinogen

P(Carcinogen): 0.286

P(NON-Carcinogen): 0.714

Reliability: the predicted compound is outside the Applicability Domain of the model

Remarks:

none

ECHA/OECD (Q)SAR's

QSAR Toolbox 3.4.0.17 [Document_1]

QSAR TOOLBOX

Input Profiling Endpoint Category Definition Data Gap Filling Report

Document Single Chemical Chemical List

New Open Close Save CAS# Name Structure Select Delete Query ChemIDs DB Inventory List

Documents

Document_1
SMILES: CC(C)CCN1c2cc(OCCCC=C)ccc2N=C1c1cccc

Filter endpoint tree...

Structure

1 [target]

Substance Identity
Physical Chemical Properties
Environmental Fate and Transport
Ecotoxicological Information
Human Health Hazards
Acute Toxicity
Bioaccumulation
Carcinogenicity
Developmental Toxicity / Teratogenicity
Genetic Toxicity
Immunotoxicity
Irritation / Corrosion
Neurotoxicity
Photoinduced Toxicity
Repeated Dose Toxicity
Sensitisation
ToxCast
Toxicity to Reproduction
Toxicokinetics, Metabolism and Distribution

CC(C)CCN1c2cc(OCCCC=C)ccc2N=C1c1cccc(CO)n1

... select filter type ... Create Apply

1 Document_1 1/0/0

12:34 26/08/2016

ECHA/OECD (Q)SAR's

QSAR Toolbox 3.4.0.17 [Document_1]

QSAR TOOLBOX

Input Profiling Endpoint Category Definition Data Gap Filling Report

Profiling Profiling Schemes

Apply New View Delete

The OECD QSAR Toolbox for Grouping Chemicals into Categories
Developed by LMC, Bulgaria

Filter endpoint tree...

1 [target]

Structure

Ionization at pH = 9
Protein binding by OASIS v1.4
Protein binding by OECD
Protein binding potency
Superfragments
Toxic hazard classification by Cramer (extension)
Toxic hazard classification by Cramer (original)
Ultimate biodeg
Endpoint Specific
Acute aquatic toxicity classification by Verhaar (M...
Acute aquatic toxicity MOA by OASIS
Aquatic toxicity classification by ECOSAR
Bioaccumulation - metabolism alerts

No pKa value
No pKb value
No alert found
No alert found
Not possible to cla...
No superfragment
High (Class III)
High (Class III)
No Data
Class 3 (unspecific...
Reactive unspecified
Benzyl Alcohols
Imidazoles
Aliphatic alcohol [...
Aromatic ether [...
Aromatic-CH2
Aromatic-H
-C=CH [alkenyl hy...
-CH- [linear]
-CH2- [linear]
Methyl [-CH3]

Profiling methods

Select All Unselect All Invert

Predefined

- ☒ Database Affiliation
- ☒ Inventory Affiliation
- ☒ OECD HPV Chemical Categories
- ☒ Substance Type
- ☒ US-EPA New Chemical Categories

General Mechanistic

- ☒ Biodeg BioHC half-life (Biowin)
- ☒ Biodeg primary (Biowin 4)
- ☒ Biodeg probability (Biowin 1)
- ☒ Biodeg probability (Biowin 2)
- ☒ Biodeg probability (Biowin 5)
- ☒ Biodeg probability (Biowin 6)
- ☒ Biodeg probability (Biowin 7)
- ☒ Biodeg ultimate (Biowin 3)
- ☒ DNA binding by OASIS v.1.4
- ☒ DNA binding by OECD
- ☒ DPRA Cysteine peptide depletion

Metabolism/Transformations

Select All Unselect All Invert

Documented

- ☐ Observed Mammalian metabolism
- ☐ Observed Microbial metabolism
- ☐ Observed Rat In vivo metabolism
- ☐ Observed Rat Liver S9 metabolism

Simulated

- ☐ Autoxidation simulator

1/0/0

12:47
26/08/2016

ECHA/OECD (Q)SAR's

Toxic hazard classification by Cramer (extension) (General Mechanistic) - Profiling Scheme Browser

Save Options Close

Properties Reference Target/Metabolites Metabolism Packet Tr Query Tree Path Report

Low (Class I)

Substances with simple chemical structures and for which efficient modes of metabolism exist, suggesting a low order of oral toxicity.

ID	Node	Result	Label
0	R1: Normal constituent of the body	Not satisfied	
2	R2: Contains functional groups associated with enhanced toxicity	Not satisfied	
4	R3: Contains elements other than C,H,O,N,divalent S	Not satisfied	
9	R7: Heterocyclic	Satisfied	
10	R8: Lactone or cyclic diester	Not satisfied	
58	R33: Has sufficient number of sulphonate or sulphamate groups	Not satisfied	High (Class III)
15	R10: 3-membered heterocycle	Not satisfied	
21	R11: Has a heterocyclic ring with complex substituent	Not satisfied	

Windows taskbar: 12:51 26/08/2016

ECHA/OECD (Q)SAR's

QSAR Toolbox 3.4.0.17 [Document_1]

QSAR TOOLBOX

Input Profiling Endpoint Category Definition Data Gap Filling Report

Categorize Delete

Define Define with metabolism Subcategorize Combine Clustering Delete Delete All

The OECD QSAR Toolbox for Grouping Chemicals into Categories
Developed by LMC, Bulgaria

Grouping methods

- Predefined
 - Database Affiliation
 - Inventory Affiliation
 - OECD HPV Chemical Categories
 - Substance Type
 - US-EPA New Chemical Categories
- General Mechanistic
 - Biodeg BioHC half-life (Biowin)
 - Biodeg primary (Biowin 4)
 - Biodeg probability (Biowin 1)
 - Biodeg probability (Biowin 2)
 - Biodeg probability (Biowin 5)
 - Biodeg probability (Biowin 6)
 - Biodeg probability (Biowin 7)
 - Biodeg ultimate (Biowin 3)
 - DNA binding by OASIS v.1.4
 - DNA binding by OECD
 - DPRA Cysteine peptide depletion
 - DPRA Lysine peptide depletion
 - Estrogen Receptor Binding
 - Hydrolysis half-life (Ka, pH 7)(Hydrowin)
 - Hydrolysis half-life (Ka, pH 8)(Hydrowin)
 - Hydrolysis half-life (Kb, pH 7)(Hydrowin)

Defined Categories

- Document_1
 - [1] Acid Chlorides <AND> Esters (Acute toxicity)
 - [21425] Not categorized (US-EPA New Chemical Categories)

Filter endpoint tree...

Structure

Substance Identity

- CAS Number
- Chemical IDs
- Chemical Name
- Molecular Formula
- Structural Formula
- Physical Chemical Properties
- Environmental Fate and Transport
- Ecotoxicological Information
- Human Health Hazards

1 [target]	2	3	4	5	6
N/A	50-14-6	50-21-5	50-32-8	50-71-5	53-59-8
NA	EINECS:2000149	EINECS:2000180	EINECS:2000285	EINECS:2000620	NA
	calciferol vitamin d2 9,10-secoergosta-... ergocalciferol (3s,5z,7e,22e)-9,1... 9,10-secoergosta-... ergocalciferol(inn) ... 9,10-secoergosta-... cyclohexanol, 4-m... (1s,3z)-[(1r,2e,4r)-... ergocalciferol (calc... cycloheanol, 4-met...	lactic acid 2-hydroxypropanoi... propanoic acid, 2-h... propanoic acid, 2-h... lactic acid [jan] 2-hydropropanoic a... lactic acid (propan...	benzo(a)pyrene benzo[a]pyrene benzo[def]chrysene benzo(a)pyrene (bap) benzo[ghi]perylene 2-hydroxypropanoic... benzopyrene 3,4-" 3,4-"	alloxane 2,4,5,6(1h,3h)-pyri... alloxan pyrimidine-2,4,5,6(...	nadp adenosi
C23H29N3O2	C28H44O	C3H6O3	C20H12	C4H2N2O4	C21H29
CC(C)CCN1c2cc(...	CC(C)C(C)C=CC(C...	CC(O)C(O)=O	c1cccc2cc3ccc4c...	O=C1C(=O)C(O)=O...	Nc1c2c...

21425 Not categorized (US-EPA New Chemical Categories)

1/0/0

13:03
26/08/2016

(Quantitative) Structural activity relationships (Q)SAR's

2. What do we want to use them for? KEY QUESTION !!!

https://echa.europa.eu/documents/10162/13655/pg_report_qsars_en.pdf

2.5 Strategy for using (Q)SAR results

In general, it is recommended to use (Q)SAR results as part of a weight of evidence (WoE) approach or as supporting information. For instance, (Q)SAR predictions can support results from tests that have not been performed according to good laboratory practice (GLP) or according to accepted guidelines, if those predictions concur with the experimental results. A compilation of several predictions with unassignable quality cannot provide an adaptation on its own.

When using (Q)SARs, it is recommended to run all (Q)SAR models available to the registrant for the endpoint to be fulfilled, especially when models are independent from each other (e.g. the algorithms are based on different descriptors, structural alerts or training sets). Agreement among predictions generated from independent and scientifically-valid (Q)SAR models increases the confidence in relying on the predictions.

Predictions that fulfil only some conditions specified in REACH Annex XI (1.3) should be disregarded or the reason for providing these predictions should be explained if it is considered that there are some benefits to providing these predictions. If the remaining (valid and adequate) predictions show small quantitative differences, the most conservative result should be chosen for further consideration. If those remaining predictions show significant quantitative differences, the registrant must decide if these differences could affect the risk assessment (for demonstrating safe use) and/or classification and labelling.

(Quantitative) Structural activity relationships (Q)SAR's

2. What do we want to use them for? KEY QUESTION !!!

How about simple physical constants for PID's?

Not available on VEGA.

Might be on ECHA toolbox (but not found so far).

(Quantitative) Structural activity relationships (Q)SAR's

2. What do we want to use them for? KEY QUESTION !!!

EPI Suite

File Edit Functions Batch Mode Show Structure Output Fugacity STP Help

EPI Suite - Welcome Screen

PhysProp Previous Get User Save User Search CAS Calculate Clear Input Fields

Draw

Input CAS #

Input Smiles: OCC1=NC(C2=NC(C=CC(OCCCC=C)=C3)=C3N2CCC(C)C)=CC=C1

Input Chem Name:

Name Lookup

Henry LC: atm-m³/mole Water Solubility: mg/L

Melting Point: Celsius Vapor Pressure: mm Hg

Boiling Point: Celsius Log Kow:

River Lake

Water Depth: 1 1 meters

Wind Velocity: 5 0.5 meters/sec

Current Velocity: 1 0.05 meters/sec

Molecular Weight: 379.51

Mol. For: C23 H29 N3 O2

Results Window

Click here for file save/print options

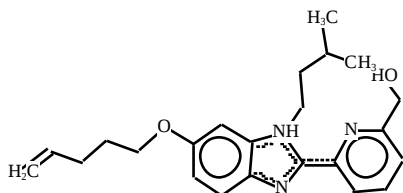
Chemical Structure:

All Results | KOWWIN | MPBPVP | Water Solubility | ECOSAR | HENRYWIN | KOAWIN | BIOWIN | BioHCwin | AEROWIN | AOPWIN | KOCWIN | HYDROWIN | BCFBAF | Volatilization | STP

Boiling Pt (deg C): 573.21 (Adapted Stein & Brown method)
Melting Pt (deg C): 246.79 (Mean or Weighted MP)
VP (mm Hg, 25 deg C): 3.99E-015 (Modified Grain method)
VP (Pa, 25 deg C): 5.32E-013 (Modified Grain method)
Subcooled liquid VP: 1.02E-012 mm Hg (25 deg C, Mod-Grain method)
: 1.36E-010 Pa (25 deg C, Mod-Grain method)

(Quantitative) Structural activity relationships (Q)SAR's

2. What do we want to use them for? KEY QUESTION !!!



SMILES : OCC1NC(C2NC(CCC(OCCCC=C)C3)C3N2CCC(C)C)CCC1

Boiling Pt, Melting Pt, Vapor Pressure Estimations (MPBPVP v1.43):

Boiling Pt (deg C): 573.21 (Adapted Stein & Brown method)

Melting Pt (deg C): 246.79 (Mean or Weighted MP)

VP(mm Hg,25 deg C): 3.99E-015 (Modified Grain method)

VP (Pa, 25 deg C) : 5.32E-013 (Modified Grain method)

Subcooled liquid VP: 1.02E-012 mm Hg (25 deg C, Mod-Grain method)

: 1.36E-010 Pa (25 deg C, Mod-Grain method)

Water Solubility Estimate from Log Kow (WSKOW v1.42):

Water Solubility at 25 deg C (mg/L): 0.1373

log Kow used: 5.24 (estimated)

(Quantitative) Structural activity relationships (Q)SAR's

4. Assimilation of findings.

VEGA quite simple – but how do we generate an operating practice without resorting to best guess by whoever is interpreting the results?

ECHA toolbox is remarkably complex and would really be better to seek training once established what we want it for as it can give more output than VEGA and also allows Data gap filling and many other functions.

ECHA toolbox is developed specifically for REACH-IT classification and reporting. It can interface with the International Uniform Chemical Information Database (ICULID) in format ready for REACH-IT registration of chemicals.

EPI suite can supply a simple (but rough) guide to many physical constants which may be of use on supply of PIDS.

