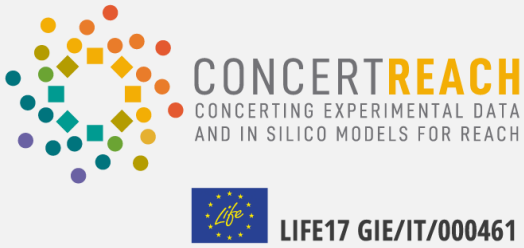
Danish (Q)SAR Database, introduction and novelty generated in the EU LIFE Concert REACH project

Eva Bay Wedebye and Nikolai Georgiev Nikolov

Workshop 27 October 2021



DTU Food
National Food Institute

CONCERTREACH
CONCERTING EXPERIMENTAL DATA AND IN SILICO MODELS FOR REACH
LIFE17 GIE/IT/000461

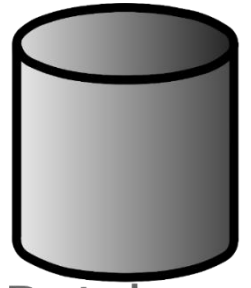
Outline

- **Danish (Q)SAR Database – what is it and examples of project we used it for**
- **Danish (Q)SAR Models, briefly**
- **DTU role in the EU LIFE Concert REACH project**

Danish (Q)SAR Database

https://qsar.food.dtu.dk/

- Developed and maintained by our DTU Food QSAR team
- Pre-generated predictions for **>650,000 mono-constituent organic substances**
- **>200 QSAR DTU/commercial/free models used** (phys.chem / HH / ENV)
- **Documentation** of models in harmonized format (QMRF)
- **Free, easy-to-use**, fast and advanced searches
- Integrated with the free **OECD QSAR Toolbox**
- >40 of the models in free DTU site: **Danish (Q)SAR Models**
- Both sites are **continuously being updated with new models**

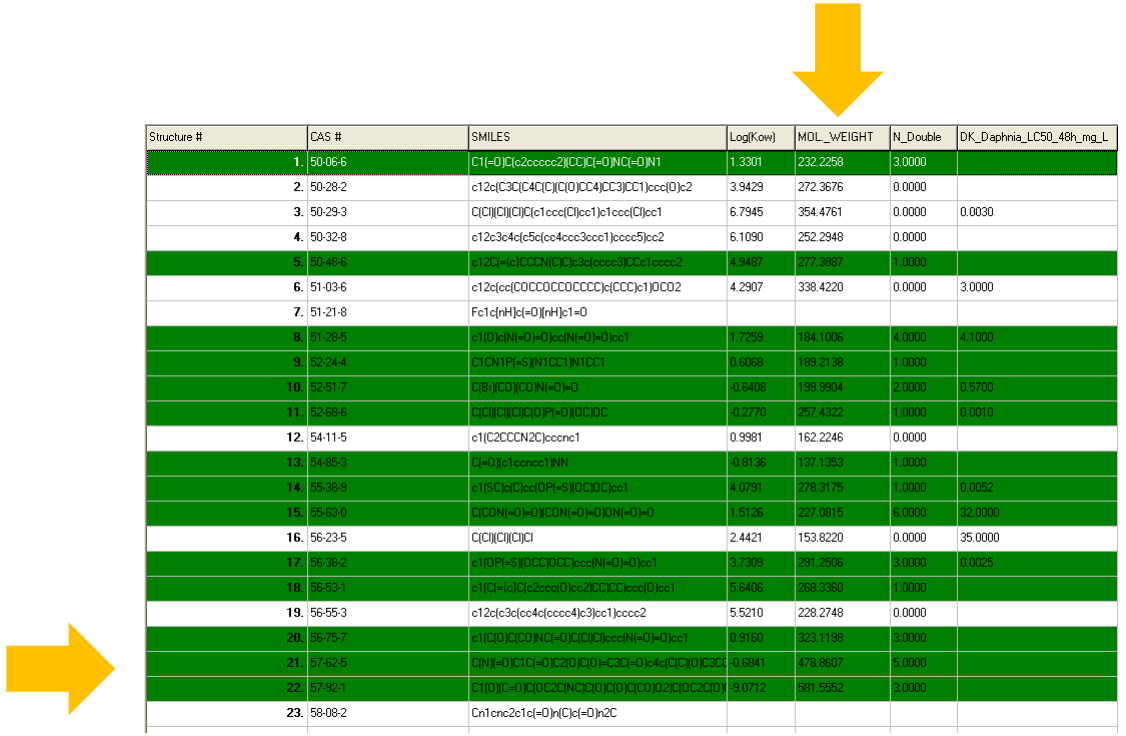


Statistics since release November 2015
>10k unique IPs made
>200k ‘real’ searches

CONCERTREACH
CONCERTING EXPERIMENTAL DATA AND IN SILICO MODELS FOR REACH
LIFE17 GIE/IT/000461

Why a predictions database

- **Quick look-up of already generated predictions** at 'top level' from many QSAR models
- **Profiling of single substances**: Integrating predictions to give a bigger picture of coherence as a step 1 in weight-of-evidence assessment
- **Screening across all contained predictions/structures**: Advanced search combinations for screening purposes
- **Building read-across cases**: Find structural analogues with the similarity function and use predictions and training set data - and/or find Tox21, PubChem etc. analogs with exp. data - to contribute to read-across justification



Structure #	CAS #	SMILES	Log(Kow)	MOL_WEIGHT	N_Double	DK_Daphnia_LC50_48h_mg_L
1.	50-06-6	C1(=O)C(c2ccccc2)(CC)C(=O)NC(=O)N1	1.3301	232.2258	3.0000	
2.	50-28-2	c12c(C3C(C4C(C)(C(O)CC4)CC3)CC1)ccc(O)c2	3.9429	272.3676	0.0000	
3.	50-29-3	C(Cl)(Cl)(Cl)C(c1ccc(Cl)cc1)c1ccc(Cl)cc1	6.7945	354.4761	0.0000	0.0030
4.	50-32-8	c12c3c4c(c5c(cc4ccc3ccc1)cccc5)cc2	6.1090	252.2948	0.0000	
5.	50-48-6	c12C(=c)CCCN(C)Cc3c(cccc3)CCc1cccc2	4.9487	277.3887	1.0000	
6.	51-03-6	c12c(cc(COCCOCCOCCCC)c(CCC)c1)OCO2	4.2907	338.4220	0.0000	3.0000
7.	51-21-8	Fc1c[nH]c(=O)[nH]c1=O				
8.	51-28-5	c1(O)c(N(=O)=O)cc(N(=O)=O)cc1	1.7259	184.1006	4.0000	4.1000
9.	52-24-4	C1CN1P(=S)(N1CC1)N1CC1	0.6068	189.2138	1.0000	
10.	52-51-7	C(Br)(CO)(CO)N(=O)=O	-0.6408	199.9904	2.0000	0.5700
11.	52-68-6	C(Cl)(Cl)(Cl)C(O)P(=O)(OC)OC	-0.2770	257.4322	1.0000	0.0010
12.	54-11-5	c1(C2CCCN2C)cccnc1	0.9981	162.2246	0.0000	
13.	54-85-3	C(=O)c1ccncc1)NN	-0.8136	137.1353	1.0000	
14.	55-38-9	c1(SC)c(C)cc(OP(=S)(OC)OC)cc1	4.0791	278.3175	1.0000	0.0052
15.	55-63-0	C(CON(=O)=O)(CON(=O)=O)ON(=O)=O	1.5126	227.0815	6.0000	32.0000
16.	56-23-5	C(Cl)(Cl)(Cl)Cl	2.4421	153.8220	0.0000	35.0000
17.	56-38-2	c1(OP(=S)(OCC)OCC)ccc(N(=O)=O)cc1	3.7309	291.2506	3.0000	0.0025
18.	56-53-1	c1(C(=c)C(c2ccc(O)cc2)CC)CC)ccc(O)cc1	5.6406	268.3360	1.0000	
19.	56-55-3	c12c(c3c(cc4c(cccc4)c3)cc1)cccc2	5.5210	228.2748	0.0000	
20.	56-75-7	c1(C(O)(C(CO)NC(=O)C(Cl)Cl)ccc(N(=O)=O)cc1	0.9160	323.1198	3.0000	
21.	57-62-5	C(N)(=O)C1C(=O)C2(O)C(O)=C3C(=O)c4c(C(C)(O)C3C1	-0.6841	478.8607	5.0000	
22.	57-92-1	C1(O)(C=O)C(OC2C(NC)C(O)C(O)C(CO)O2)C(OC2C(O)	-9.0712	581.5552	3.0000	
23.	58-08-2	Cn1cnc2c1c(=O)n(C)c(=O)n2C				

(Q)SAR software/models

Commercial and free software/models with use / publication conditions from:

MultiCASE Incorporated

Leadscope®

SciMatics From Disorder to Discovery

ACD/Labs

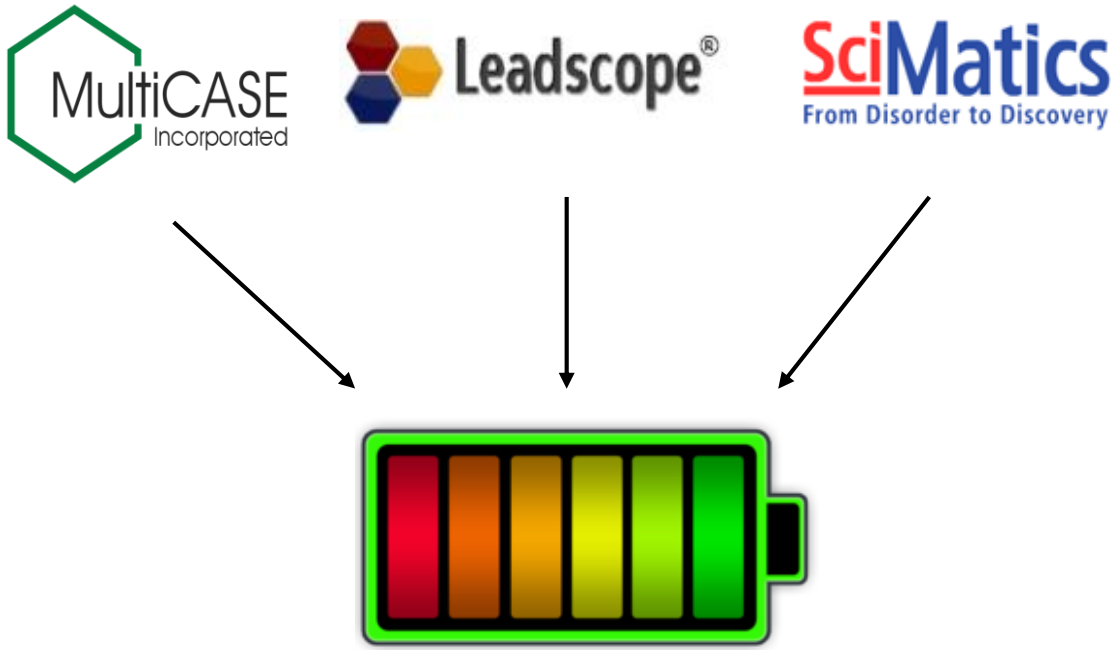
EPI

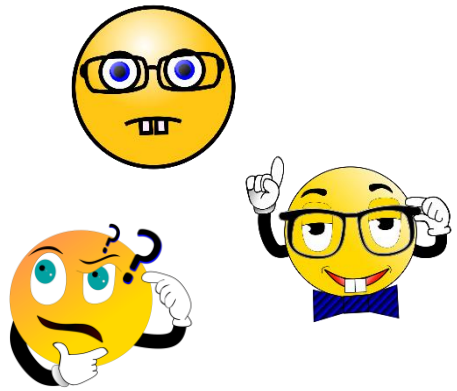
Laboratory of Mathematical Chemistry LMC Oasis

QSAR TOOLBOX

VEGAHUB

Development of battery approach for many training sets





<u>Majority vote for a substance:</u>
requiring models from at least **two systems** to be **within AD and agree** on prediction

Pictures from https://itunes.apple.com/us/app/battery-health-monitor-stats, https://pixabay.com/da/smiley-nørd-briller-pc-ekspert-1914523/, http://colegio.etep.edu.br/noticia/como-falar-sobre-os-meus-pontos-fracos/, https://www.onlinelabels.com/clip-art/Emoticons%3A-Nerd-face-108553.htm

Danish (Q)SAR Database DTU Ames battery model

- A battery prediction of Positive or Negative inside Applicability Domain (AD), requires predictions in at least 2 of the following 3 models to agree and be inside AD:

Endpoint	N (training set)	Software	Cross validation 5 * 2-fold (%)
Bacterial reverse mutation test (Ames test in S. typhimurium, in vitro)	**4,102*** (same training set modelled in 3 software systems)	CASE Ultra	Sens=83.9, Spec=89.1, BA=86.5
		Leadscope	Sens=84.3, Spec=85.7, Conc=85.0
		SciQSAR	Sens=79.3, Spec=79.1, Conc=79.2

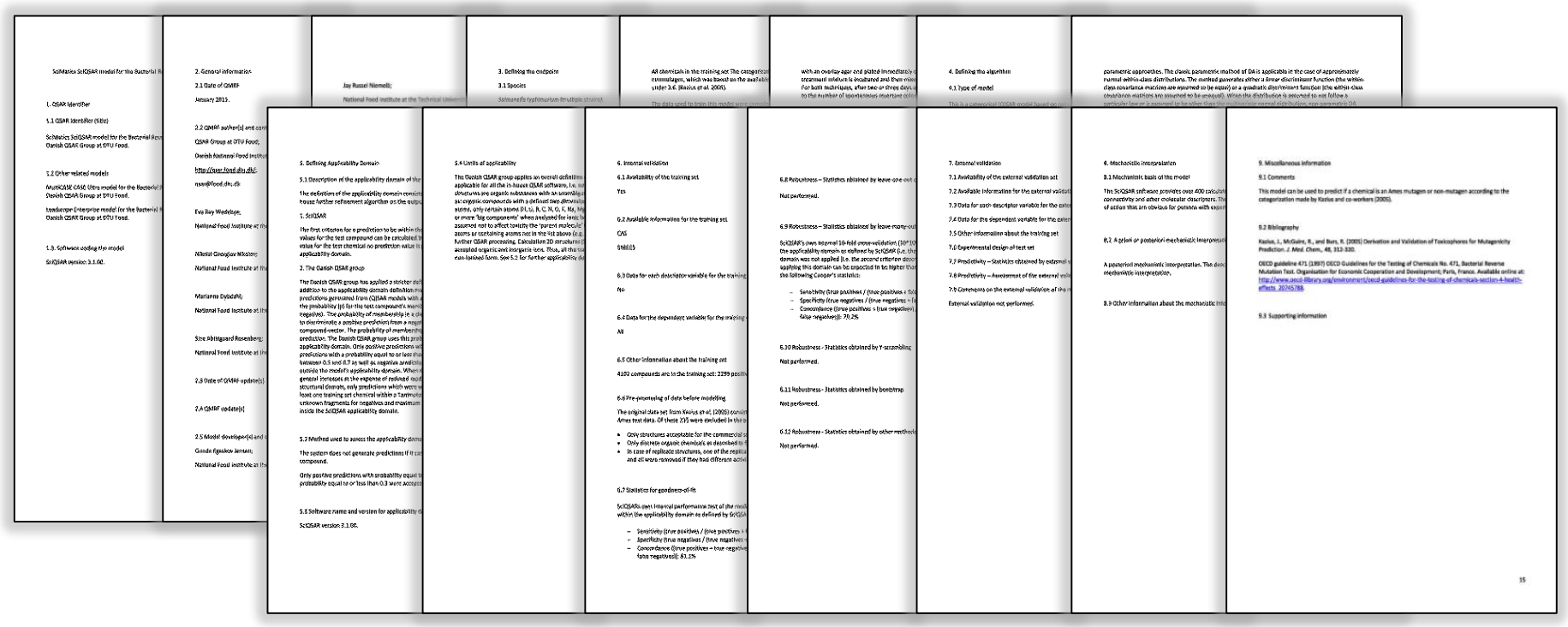
- A quick **external validation** (not published) with Hansen et al.** substances, which were external to the DTU Ames battery model gave:
  - Sensitivity: 523/(523+120)=81.3%
  - Specificity: 441/(441+53)=89.3%
  - Balanced Accuracy (BA) = 85.3%

* Kazius et al., J. Med. Chem. 2005, 48, p. 312-320

** Hansen et al., J. Chem. Inf. Model. 2009, 49, p. 2077–2081, NB only substances included in the Danish DB were used

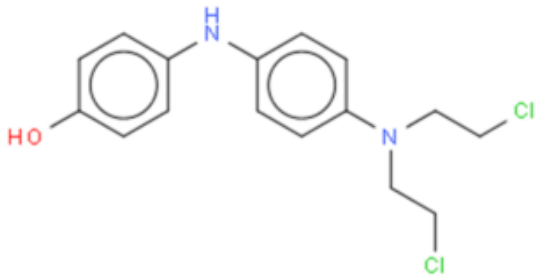
Validation and documentation

- QSAR Model Reporting Formats (QMRFs) for >130 models



(Q)SAR predicted profile

Structure (as used for QSAR prediction):



SMILES (used for QSAR prediction): c1(Nc2ccc(O)cc2)ccc(N(CCCl)CCCl)cc1

ID

REACH EC Number (pre-registration, by 2013)		REACH EC Number (registration, by Dec. 2019)	
Registry Number	63979-55-5	PubChem CID	
EU CLP Harmonized Classification*		DK-EPA / DTU QSAR-based CLP Advisory Classification	
REACH registration cumulated minimum annual tonnage		US TSCA (Oct. 2021)	
Tox21 (2019)		ToxCast (Oct. 2021)	
Molecular Formula	C16 H18 CL2 N2 O1	Molecular weight (g/mole)	325.24
Chemical Name	Diphenylamine, 4'-(bis(2"-chloroethyl)amino)-4-hydroxy-		

(Annex VI to CLP up to and including the 9th ATP, and including Nordic Council of Minister SPIN list for group entries)

Melting point, Boiling point and Vapour pressure

Melting Point (deg C)	181.68	Melting Point Experimental (deg C)	
Boiling Point (deg C)	447.03	Boiling Point Experimental (deg C)	
Vapour Pressure (atm)	EPI.Estimated_VP_atm	Vapour Pressure Experimental (atm)	EPI.Exp_VP_atm
Vapour Pressure (mm Hg)	2.29E-009	Vapour Pressure Experimental (mm Hg)	
Vapour Pressure (Pa)	3.053E-007	Vapour pressure Subcooled Liquid (Pa)	1.31E-005

EPI MPBPVP models

Henry's Law Constant

HLC Bond Method (atm-m3/mole)	3.8E-013	HLC Group Method (atm-m3/mole)	
HLC Via VP/WSol (atm-m3/mole)	1.113E-010	HLC Via VP/WSol (Pa-m3/mole)	1.128E-005
Henrys Law Const. Exp db (Pa-m3/mole)		Henrys Law Const. Exp db (atm-m3/mole)	

EPI HENRYWIN models

Water Solubility

Water solubility from Kow (mg/L)	8.805	Water solubility from Fragments (mg/L)	4.4439
Water solubility Exp (mg/L)		Water solubility Exp Ref	

EPI WATERNT model

Hydrolysis

Hydrolysis Ka half-life pH 7	Hydrolysis Kb half-life pH 7
Hydrolysis Ka half-life pH 8	Hydrolysis Kb half-life pH 8

EPI HYDROWIN model

pKa

pKa Acid	10.4
- Standard deviation (±)	0.8
pKa Base	3.5
- Standard deviation (±)	0.6

ACDLabs model

pKa estimate 999: no acidic moiety found. pKa estimate -999: no basic moiety found.

Partition coefficients

pH	1	4	5	6	7	8	9
LogD	1.74	3.86	3.97	3.99	3.99	3.99	3.97

Minimum LogD in the pH interval 4-9	3.86	Maximum LogD in the pH interval 4-9	3.99

ACDLabs models

LogD: Log octanol-water partition coefficient, which for ionizable compounds varies with the pH-dependent amounts of neutral and ionized species

Log Koa	14.319	Log Kaw	-10.809

EPI KOAWIN models

Koa: octanol-air partition coefficient. Kaw: air-water partition coefficient.

Log Kow	3.51		
Log Kow Exp		Log Kow Exp Ref	

EPI WSKOW model

LogKow: log octanol-water partition coefficient

Kp (m3/ug) Mackay-based	0.23	Kp (m3/ug) Koa-based	51.2
Phi Junge-Pankow-based	0.892	Phi Mackay-based	0.948
Phi Koa-based	1		

EPI AEROWIN models

Kp: particle-gas partition coefficient. Phi: fraction of substance sorbed to atmospheric particulates

Koc from MCI (L/kg)	23040	Log Koc from MCI	4.3624
Koc from Kow (L/kg)	884.4	Log Koc from Kow	2.9466

EPI KOCWIN models

Koc: soil adsorption coefficient of organic compounds. Kow: octanol-water partition coefficient. MCI: first order Molecular Connectivity Index

Level III Fugacity Environmental Partitioning, emission to air, water and soil

	Air	Water	Soil	Sediment
Mass Amount (%)	1.92E-005	7.02	80.7	12.3
Half-Life (hr)	1.24	1440	2880	13000
Emissions (kg/hr)	1000	1000	1000	0

EPI Level III Fugacity Model

Persistence time (hr)	3250
Persistence time (days)	135.4167

EPI Level III Fugacity Model

Level III Fugacity Environmental Partitioning, emission only to water

	Air	Water	Soil	Sediment
Mass Amount (%)	2.56E-013	36.3	5.09E-007	63.7
Half-Life (hr)	1.24	1440	2880	13000
Emissions (kg/hr)	0	1000	0	0

EPI Level III Fugacity Model

Persistence time (hr)	1710
Persistence time (days)	71.25

EPI Level III Fugacity Model

Sewage Treatment Plant (STP) overall chemical mass balance using 10,000 hr

	Total removal	Biodegradation	Sludge Adsorption	Volatilization
(%)	13.26	0.19	13.08	0

EPI STPWIN model

Atmospheric oxidation (25 deg C)

	OH	Ozone
Half-Life (d)	0.0516	0
Half-Life (hr)	0.619	
Overall Rate Const. (OH: E-12 cm3/molecule-sec and OZ: E-17 cm3/molecule-sec)	207.2672	

EPI AOPWIN models

Biodegradation

Biowin1 (linear model) Probability of Rapid Biodegradation	0.0467
Biowin2 (non-linear model) Probability of Rapid Biodegradation	0.0002
Biowin3 Expert Survey Ultimate Biodegradation	1.8007
Biowin3 Expert Survey Ultimate Timeframe	months
Biowin4 Expert Survey Primary Biodegradation	2.8206
Biowin4 Exp. Survey Primary Timeframe	weeks
Biowin5 (MITI linear model) Biodegradation Probability	-0.1682
Biowin6 (MITI non-linear model) Biodegradation Probability	0.0009
Biowin7 (Anaerobic Linear) Biodegradation Probability	-1.1247
Petroleum Hydrocarbon Biodegradation Half-Life (days)	

EPI BIOWIN models

SkinBiowin1 and Biowin2: ≥0.5: "Rapid" <0.5: "Slow"
Biowin3 and Biowin4: 5 ~ hours; 4 ~ days; 3 ~ weeks; 2 ~ months; 1 ~ years.
Biowin5 and Biowin6: ≥0.5: "Readily", <0.5: "Not readily".
Biowin7: ≥0.5: "Fast", <0.5: "Slow"

	Exp	Battery	CASE Ultra	Leadscope	SciQSAR
Not Ready Biodegradability (POS=Not Ready)		POS_IN	POS_IN	POS_IN	NEG_OUT

DTU-developed models

Bioaccumulation

BCF (L/kg wet-wt)	95.83
Log BCF (L/kg wet-wt)	1.982
Whole Body Primary Biotransformation Fish Half-Life (days)	0.1804
BCF Arnot-Gobas (upper trophic) Including Biotransformation (L/kg wet-wt)	62.25
BCF Arnot-Gobas (upper trophic) Zero Biotransformation (L/kg wet-wt)	341.5
BAF Arnot-Gobas (upper trophic) Including Biotransformation (L/kg wet-wt)	62.25
BAF Arnot-Gobas (upper trophic) Zero Biotransformation (L/kg wet-wt)	485.2

EPI BCFBAF models

BCF: Bioconcentration factor, BAF: Bioaccumulation factor

Aquatic toxicity

	Exp	Battery	Leadscope	SciQSAR
Fathead minnow 96h LC50 (mg/L)			0.6731269	0.7366208
Domain		OUT	OUT	OUT
Daphnia magna 48h EC50 (mg/L)		0.4565404	0.3870346	0.5260463
Domain		IN	IN	IN
Pseudokirchneriella s. 72h EC50 (mg/L)			0.4397876	0.04881282
Domain		OUT	OUT	OUT

DTU-developed models

	Fish 96h	Daphnid 48h	Green Algae 96h
LC50 (Fish) or EC50 (Daphnid and Algae) for Most Toxic Class (mg/L)	4.035	2.378	9.446
Max. Log Kow for Most Toxic Class	7	7	7
Most Toxic Class	Phenols	Phenols	Phenols
Note			Chemical may not be soluble enough

EPI ECOSAR models

ECOSAR Classes: Phenols

Oral absorption

Lipinski's Rule-of-five score (bioavailability)	0
Absorption from gastrointestinal tract for 1 mg dose (%)	100
Absorption from gastrointestinal tract for 1000 mg dose (%)	50

Leadscope model on Lipinski's Rule-of-five. Equation from literature on GI abs.

Lipinski scores of 0 or 1: The substance may be bioavailable. Lipinski scores of 2, 3 or 4: The substance may not be bioavailable.

Skin absorption

Dermal absorption (mg/cm2/event)	0.000248

EPI DERMWIN model

Brain/blood Distribution

Log brain/blood partition coefficient	0.411

Equation from literature

Partitioning between the two tissues at equilibrium. >1: high, >0 to <1: medium, >-1 to <0, fair, <-1: low.

Metabolism

	Exp	Battery	CASE Ultra	Leadscope	SciQSAR
CYP2C9 substrates (Human clinical data)		NEG_IN	NEG_IN	NEG_IN	NEG_IN
CYP2D6 substrates (Human clinical data)		INC_OUT	POS_IN	INC_OUT	NEG_IN

DTU-developed models

Acute toxicity in Rodents

	LD50 (mg/kg/d)	Reliability Index
Rat Oral	45.29	0.46
Rat Intraperitoneal	15.89	0.33
Mouse Oral	57.58	0.55
Mouse Intraperitoneal	84.67	0.32
Mouse Intravenous	12.12	0.27
Mouse Subcutaneous	7.33	0.46

ACDLabs models

Reliability index: <0.3 = Not reliable prediction quality; 0.3-0.5 = borderline prediction quality; 0.5-0.75 = moderate prediction quality; >0.75 = high prediction quality.

MRDD in Humans

	Exp	Battery	CASE Ultra	Leadscope	SciQSAR
MRDD in Humans ≤ 2.69 mg/kg-bw/d		POS_IN	POS_OUT	POS_IN	POS_IN

DTU-developed models

Model based on data on pharmaceuticals. Maximum recommended daily dose in pharmaceutical clinical trials employing primarily oral route of exposure and daily treatments, usually for 3-12 months.

Irritation and Sensitization

	Exp	Battery	CASE Ultra	Leadscope	SciQSAR
Severe Skin Irritation in Rabbit		POS_IN	POS_IN	POS_IN	NEG_IN
Allergic Contact Dermatitis in Guinea Pig and Human*	N/A	POS_IN	POS_IN	POS_IN	POS_IN
Respiratory Sensitisation in Humans		INC_OUT	INC_OUT	POS_OUT	NEG_OUT

DTU-developed models

**Based on commercial training set*

	VEGA	ADI
Skin Sensitization (CAESAR)	POS_Low	0.444

CAESAR skin sensitization model is version 2.1.6 contained in VEGA command line version 1.1.2 BETA 5 with calculation core version 1.2.4

Prediction: POS = Sensitizer, NEG = Non-sensitizer, SUSP.POS = Suspected sensitizer, POSS.NEG = Possible Non-sensitizer, Exp = experimental value, Good = Good reliability, Mod = Moderate reliability, Low = Low reliability.

Protein binding by OASIS, alerts in:	
- parent only	Alkyl halides
- metabolites from skin metabolism simulator only	Aldehydes; Alkyl halides; alpha-Activated haloalkanes; Mustard compounds; Quinone methide(s)/imines, Quinoide oxime structure, Nitroquinones, Naphthoquinone(s)/imines
- metabolites from auto-oxidation simulator only	Alkyl halides; Quinone methide(s)/imines, Quinoide oxime structure, Nitroquinones, Naphthoquinone(s)/imines
Protein binding by OECD, alerts in:	
- parent only	Mustards
- metabolites from skin metabolism simulator only	alpha-Halocarbonyls; Mono-carbonyls; Mustards; Polarised alkene - ketones; Quinone-imine
- metabolites from auto-oxidation simulator only	Mustards; Polarised alkene - ketones; Quinone-imine
Protein binding potency Cys (DRPA 13%), alerts in:	
- parent only	Out of mechanistic domain
- metabolites from skin metabolism simulator only	DPRA above 21% (DPRA 13%) >> Non-Conjugated monoaldehydes (reactive); DPRA above 21% (DPRA 13%) >> p-Phenylenediamine derivatives

- metabolites from auto-oxidation simulator only	DPRA above 21% (DPRA 13%) >> p-Phenylenediamine derivatives
Protein binding potency Lys (DRPA 13%), alerts in:	
- parent only	Out of mechanistic domain
- metabolites from skin metabolism simulator only	DPRA above 21% (DPRA 13%) >> Aminophenol derivatives (reactive); DPRA less than 9% (DPRA 13%) >> Non-alpha,beta-conjugated monoaldehydes (non reactive)
- metabolites from auto-oxidation simulator only	DPRA above 21% (DPRA 13%) >> Aminophenol derivatives (reactive)
Keratinocyte gene expression, alerts in:	
- parent only	Very high gene expression >> Substituted para- and ortho-phenylenediamines, aminophenols and benzenediols
- metabolites from skin metabolism simulator only	High gene expression >> Non-conjugated aldehydes and dialdehydes; Moderate gene expression >> Fragrance aldehydes; Very high gene expression >> alpha, beta-Unsaturated carbonyl compounds; Very high gene expression >> Substituted para- and ortho-phenylenediamines, aminophenols and benzenediols
- metabolites from auto-oxidation simulator only	Very high gene expression >> alpha, beta-Unsaturated carbonyl compounds
Protein binding potency GSH, alerts in:	
- parent only	Not possible to classify according to these rules (GSH)

OECD QSAR Toolbox v.4.1 profilers

Profiler predictions are supporting information to be used together with the relevant QSAR predictions

Endocrine and Molecular Endpoints

	Exp	Battery	CASE Ultra	Leadscope	SciQSAR
Estrogen Receptor α Binding, Full training set (Human *in vitro*)		INC_OUT	NEG_IN	INC_OUT	POS_IN
Estrogen Receptor α Binding, Balanced Training Set (Human *in vitro*)		POS_IN	POS_IN	POS_IN	NEG_IN
Estrogen Receptor α Activation (Human *in vitro*)		POS_OUT	INC_OUT	INC_OUT	POS_IN
Estrogen Receptor Activation, CERAPP data (*in vitro*)		N/A	N/A	NEG_IN	N/A
Androgen Receptor Inhibition (Human *in vitro*)		INC_OUT	INC_OUT	NEG_IN	POS_IN
Androgen Receptor Binding, CoMPARA data (*in vitro*)		N/A	N/A	POS_OUT	N/A
Androgen Receptor Inhibition, CoMPARA data (*in vitro*)		N/A	N/A	POS_IN	N/A
Androgen Receptor Activation, CoMPARA data (*in vitro*)		N/A	N/A	NEG_IN	N/A
Thyroperoxidase (TPO) inhibition QSAR1 (Rat *in vitro*)		N/A	N/A	POS_IN	N/A
Thyroperoxidase (TPO) inhibition QSAR2 (Rat *in vitro*)		N/A	N/A	POS_IN	N/A
Thyroid Receptor α Binding (Human *in vitro*)					
- mg/L			52024.02	201.7734	69.71142
- μM			159955.8	620.3831	214.3384
- Positive for $IC_{50} \leq 10$ μM					
- Positive for $IC_{50} \leq 100$ μM					
- Domain		OUT	OUT	OUT	OUT
Thyroid Receptor β Binding (Human *in vitro*)					
- mg/L			10524.57	6.459723	628.4979
- μM			32359.38	19.86141	1932.413
- Positive for $IC_{50} \leq 10$ μM					
- Positive for $IC_{50} \leq 100$ μM					
- Domain		OUT	OUT	OUT	OUT
Arylhydrocarbon (AhR) Activation – Rational final model (Human *in vitro*)		N/A	N/A	NEG_IN	N/A
Arylhydrocarbon (AhR) Activation – Random final model (Human *in vitro*)		N/A	N/A	NEG_OUT	N/A
Pregnane X Receptor (PXR) Binding (Human *in vitro*)	N/A	INC_OUT	POS_OUT	POS_OUT	INC_OUT
Pregnane X Receptor (PXR) Binding (Human *in vitro*) NEW		N/A	N/A	NEG_IN	N/A
Pregnane X Receptor (PXR) Activation (Human *in vitro*)		N/A	N/A	INC_OUT	N/A
Pregnane X Receptor (PXR) Activation (Rat *in vitro*)		N/A	N/A	NEG_IN	N/A
CYP3A4 Induction (Human *in vitro*)		N/A	N/A	NEG_OUT	N/A
Constitutive Androstane Receptor (CAR) Activation at max. 20 μM (*in vitro*)		N/A	N/A	NEG_IN	N/A
Constitutive Androstane Receptor (CAR) Activation at max. 50 μM (*in vitro*)		N/A	N/A	POS_IN	N/A
Constitutive Androstane Receptor (CAR) Inhibition at max. 20 μM (*in vitro*)		N/A	N/A	POS_IN	N/A
Constitutive Androstane Receptor (CAR) Inhibition at max. 50 μM (*in vitro*)		N/A	N/A	POS_IN	N/A

DTU-developed models

Estrogen Receptor Binding, alerts in:	
- parent only	Strong binder, OH group
- metabolites from *in vivo* Rat metabolism simulator only	Strong binder, NH2 group; Strong binder, OH group; Moderate binder, NH2 group; Weak binder, OH group
- metabolites from Rat liver S9 metabolism simulator only	Strong binder, OH group

rtER Expert System - USEPA, alerts in:	
- parent only	No alert found
- metabolites from *in vivo* Rat metabolism simulator only	No alert found
- metabolites from Rat liver S9 metabolism simulator only	No alert found

OECD QSAR Toolbox v.4.2 profilers

Profiler predictions are supporting information to be used together with the relevant QSAR predictions

Developmental Toxicity

	Battery	CASE Ultra	Leadscope	SciQSAR
Teratogenic Potential in Humans	POS_IN	POS_IN	POS_IN	NEG_IN

DTU-developed models based on commercial training set

Genotoxicity - Structural Alerts for DNA Reactivity

	Battery	CASE Ultra	Leadscope	SciQSAR
Ashby Structural Alerts	POS_IN	POS_IN	POS_IN	INC_OUT

DTU-developed models based on commercial training set

DNA binding by OASIS, alerts in:	
- parent only	Haloalkanes Containing Heteroatom; Nitrogen and Sulfur Mustards
DNA binding by OECD, alerts in:	
- parent only	Mustards; Tertiary aromatic amine

OECD QSAR Toolbox v.4.2 profilers

Profiler predictions are supporting information to be used together with the relevant QSAR predictions

In vitro Genotoxicity - Bacterial Reverse Mutation Test (Ames test)

	Exp	Battery	CASE Ultra	Leadscope	SciQSAR
Ames test in S. typhimurium (*in vitro*)		POS_IN	POS_IN	POS_IN	POS_IN
*Direct Acting Mutagens (without S9)	N/A	NEG_OUT	POS_OUT	NEG_IN	NEG_OUT
*Base-Pair Ames Mutagens	N/A	INC_OUT	INC_OUT	NEG_IN	POS_IN
*Frameshift Ames Mutagens	N/A	POS_IN	NEG_IN	POS_IN	POS_IN
*Potent Ames Mutagens, Reversions ≥ 10 Times Controls	N/A	POS_IN	POS_IN	POS_IN	POS_IN

DTU-developed models

* The four models (Direct Acting mutagens (without S9), Base-Pair Ames Mutagens, Frameshift Ames Mutagens, Potent Ames Mutagens) should not be used to determine if substances are Ames mutagens, but can be used for indication of mechanism or potency for cases where the main Ames model (Ames test in S. typhimurium (*in vitro*)) is POS_IN.

	VEGA	Mut. / Non-mut. scores	Used models
Mutagenicity consensus	POS	0.68 / 0	4

Mutagenicity (Ames) consensus model version 1.0.2 contained in VEGA version 1.1.4 with calculation core version 1.2.4

Prediction: POS = Mutagenic, NEG = Non-mutagenic.

VEGA

ISS	CAESAR	SarPy	KNN
POS_Good	POS_Mod	POS_Mod	POS_Mod

Four individual models in mutagenicity consensus model version 1.0.2 contained in VEGA version 1.1.4 with calculation core version 1.2.4

Prediction: POS = Mutagenic, NEG = Non-mutagenic, SUSP.POS = Suspected mutagenic, POSS.NEG = Possible Non-mutagenic, Exp = experimental value, Good = Good reliability, Mod = Moderate reliability, Low = Low reliability.

DNA alerts for AMES by OASIS, alerts in:	
- parent only	No alert found
In vitro mutagenicity (Ames test) alerts by ISS, alerts in:	
- parent only	S or N mustard

OECD QSAR Toolbox v.4.2 profilers

Profiler predictions are supporting information to be used together with the relevant QSAR predictions

Other *in vitro* Genotoxicity Endpoints

	Exp	Battery	CASE Ultra	Leadscope	SciQSAR
Chromosome Aberrations in Chinese Hamster Ovary (CHO) Cells*	N/A	NEG_OUT	INC_OUT	INC_OUT	NEG_IN
Chromosome Aberrations in Chinese Hamster Lung (CHL) Cells		POS_OUT	POS_OUT	POS_IN	INC_OUT
Mutations in Thymidine Kinase Locus in Mouse Lymphoma Cells		POS_IN	POS_IN	POS_IN	POS_IN
Mutations in HGPRT Locus in Chinese Hamster Ovary (CHO) Cells		POS_OUT	POS_OUT	POS_IN	NEG_OUT
Unscheduled DNA Synthesis (UDS) in Rat Hepatocytes		NEG_IN	NEG_IN	NEG_IN	NEG_IN
Syrian Hamster Embryo (SHE) Cell Transformation		POS_OUT	POS_OUT	INC_OUT	POS_IN

DTU-developed models
**Based on commercial training set*

HGPRT: Hypoxanthine-guanine phosphoribosyltransferase

DNA alerts for CA and MNT by OASIS, alerts in:	
- parent only	No alert found
Protein binding alerts for Chromosomal aberration by OASIS, alerts in:	
- parent only	Nitrogen Mustard

OECD QSAR Toolbox v.4.2 profilers
CA: Chromosomal aberration, MNT: Micronucleus test
Profiler predictions are supporting information to be used together with the relevant QSAR predictions

In vivo Genotoxicity Endpoints

	Exp	Battery	CASE Ultra	Leadscope	SciQSAR
Sex-Linked Recessive Lethal (SLRL) Test in Drosophila m.		POS_IN	POS_IN	POS_IN	POS_IN
Micronucleus Test in Mouse Erythrocytes		POS_IN	POS_OUT	POS_IN	POS_IN
Dominant Lethal Mutations in Rodents		POS_IN	POS_IN	POS_IN	POS_IN
Sister Chromatid Exchange in Mouse Bone Marrow Cells		POS_IN	INC_OUT	POS_IN	POS_IN
Comet Assay in Mouse		POS_IN	POS_IN	NEG_OUT	POS_IN

DTU-developed models

In vivo mutagenicity (Micronucleus) alerts by ISS, alerts in:	
- parent only	S or N mustard

OECD QSAR Toolbox v.4.2 profilers

Profiler predictions are supporting information to be used together with the relevant QSAR predictions

Carcinogenicity

	E Ultra	Leadscope
FDA RCA Cancer Male Rat	POS_OUT	NEG_IN
FDA RCA Cancer Female Rat	POS_IN	POS_IN
FDA RCA Cancer Rat	POS_OUT	POS_IN
FDA RCA Cancer Male Mouse	POS_IN	POS_IN
FDA RCA Cancer Female Mouse	POS_IN	POS_IN
FDA RCA Cancer Mouse	POS_IN	POS_IN
FDA RCA Cancer Rodent	POS_IN	POS_IN

Commercial models from CASE Ultra and Leadscope

FDA RCA: Data from US Food and Drug Administration as part of Research Cooperation Agreement

Carcinogenicity (genotox and nongenotox) alerts by ISS, alerts in:	
- parent only	S or N mustard (Genotox); Structural alert for genotoxic carcinogenicity
Oncologic Primary Classification, alerts in:	
- parent only	Aromatic Amine Type Compounds; Nitrogen Mustards Reactive Functional Groups; Phenol Type Compounds

OECD QSAR Toolbox v.4.2 profilers

Profiler predictions are supporting information to be used together with the relevant QSAR predictions

	Exp	Battery	CASE Ultra	Leadscope	SciQSAR
Liver Specific Cancer in Rat or Mouse		NEG_OUT	NEG_IN	INC_OUT	INC_OUT

DTU-developed models

Abbreviations

INC: inconclusive. A definite call within the defined applicability domain could not be made.

NEG: negative

POS: positive

IN: inside applicability domain

OUT: outside applicability domain

Exp: Experimental values, from EpiSuite experimental databases or DK DTU QSAR models training sets.

N/A: Not applicable, either because training set data cannot be released for commercial or proprietary models / training sets, or because the model was not developed in a given QSAR software (i.e. a given prediction is not available as the model version does not exist).

Important notes

This is an automatically generated report from the Danish (Q)SAR Database, http://qsar.food.dtu.dk.

For predictions from CASE Ultra, Leadscope, SciQSAR as well as the Acute toxicity in rodent from ACDLabs information on the software versions can be found in the QMRFs. For the other predicted properties the software versions are:

EPI MPBPWIN v1.43

EPI HENRYWIN v3.20

EPI WSKOW v1.42

EPI WATERNT v1.01

EPI KOAWIN v1.10

EPI AEROWIN v1.00

EPI KOCWIN v2.00

EPI Level III Fugacity Model (EPI Suite v4.11)

EPI STPWIN (EPI Suite v4.11)

EPI AOPWIN v1.92

EPI BIOWIN v4.10

EPI BCFBAF v3.01

EPI ECOSAR v1.11

EPI DERMWIN v2.02

ACD/ ToxSuite 2.95.1 Ionization\pKa

ACD/ ToxSuite 2.95.1 Ionization\ LogD

ACD/ ToxSuite 2.95.1

It is recommended to run the latest version of the EPI Suite Programs in preference of the predictions given in this document when these endpoints are of importance and new versions have been released from the United States Environmental Protection Agency in comparisons. EPI Suite can be downloaded from the US EPA homepage: http://www.epa.gov/oppt/exposure/pubs/episuitedl.htm

For further information on the applied systems, see the following homepages:

Case Ultra: http://www.multicase.com/case-ultra

Leadscope: http://www.leadscope.com/

SciQSAR: http://lhasa-llc.com/

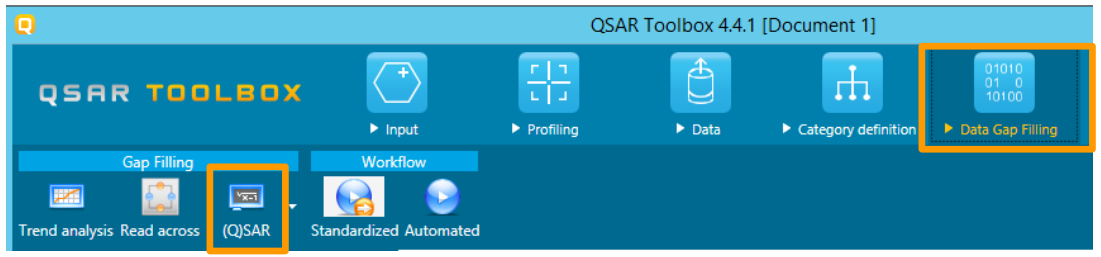
ToxSuite: http://www.acdlabs.com/

Copyright notice, terms and conditions of use

Imbedded film not included in this pdf..

Integration with the OECD QSAR Toolbox in 2018

- Previous Danish (Q)SAR Database **incorporated**
- Danish (Q)SAR Database **integrated** via on-the-fly-access



CONCERTREACH
CONCERTING EXPERIMENTAL DATA AND IN SILICO MODELS FOR REACH

LIFE17 GIE/IT/000461

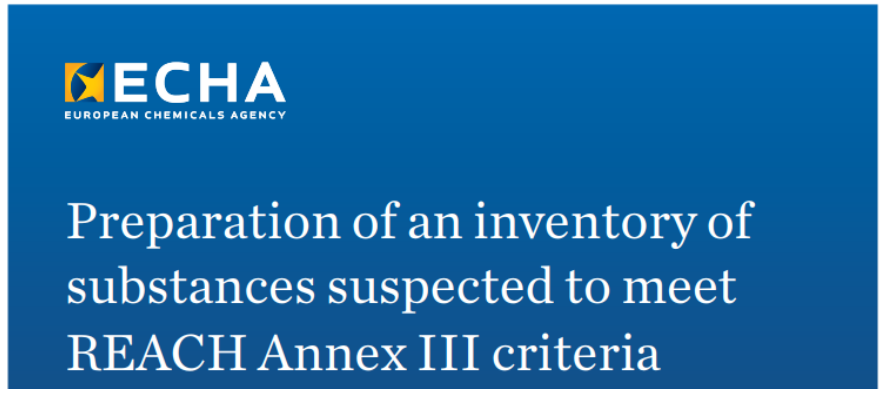
REACH Annex III 2016

Structures and predictions from the Danish (Q)SAR Database were used together with predictions from other systems for the ECHA inventory of substances suspected to meet REACH Annex III criteria

Substances registering in the tonnage band of 1–10 tonnes/year meeting either one or both of the Annex III criteria have to **provide full Annex VII information**:

a. substances predicted (i.e. by the use of QSARs or other evidence) to likely meet criteria for **CMR** category 1A or 1B or Annex XIII criteria (i.e. **PBT** and **vPvB**);
b. substances with **dispersive or diffuse use(s)** AND predicted to likely meet criteria for **any health or environmental hazard** classes or differentiations under CLP Regulation.

The REACH **registration deadline** for phase-in substances in the 1-10 tonnage band was **31 May 2018**



Use of the Danish (Q)SAR Database (incl. previous versions), examples

- **DK EPA advisory classifications** screenings (2016-18, 2010, 2009, 2001)
- Grouping and **category approach for brominated flame retardants** for DK-EPA (2016)
- **Endocrine activity screening** for DK EPA screening 72,000 REACH substances (2014)
- DK EPA screening 72,000 substances identifying **potential CMR substances** of relevance under the REACH regulation (2013)
- **EU FP7 ChemScreen**, WP on QSAR pre-screen for **reproductive toxicity** screening 72,000 REACH substances (2010-2013)
- **PMT screening** (2017-20), PBT screening (2002), POP screening (1999)

- **EU REACH activities** – REACH implementation projects (RIPs), substance evaluations, dossier evaluations and screening for SEv-candidate substances (for clarification whether SVHC nomination would be relevant), commenting of (Q)SAR-related guidances (e.g. RAAF).
- **OECD activities** – development of (Q)SAR validation principles (Setubal/OECD), QSAR Toolbox, commenting of guidance documents, contributing QSAR predictions concerning mono-constituent organic SIDS chemicals to former HPV programme for >10 y from 2000 etc.

Danish (Q)SAR Models

powered by Leadscope Predictive Data Miner

Ministry of Environment and Food of Denmark

Leadscope®

DTU Food National Food Institute

DTU

Home | New query | Quick start guide | Model documentation | Contact | Danish (Q)SAR Database

Molecule Id (optional):

Select models

- Select all
- Environmental
- ADME
- Endocrine/molecular
- Endocrine/molecular 2
- Genotoxicity/cancer
- Other endpoints

- Estrogen receptor (ER)
  - ER alpha binding, all (human in vitro)
  - ER alpha binding, balanced (human in vitro)
  - ER alpha activation (human in vitro)
  - ER Activation (in vitro, CERAPP data)
- Androgen receptor (AR)
  - AR inhibition (human in vitro)
  - AR binding (in vitro, CoMPARA data)
  - AR activation (in vitro, CoMPARA data)
  - AR inhibition (in vitro, CoMPARA data)
- Thyroid-related endpoints
  - Thyroperoxidase (TPO) inhibition QSAR1 (in vitro)
  - Thyroperoxidase (TPO) inhibition QSAR2 (in vitro)

Input structure

C N O F P S H

Predict

Currently holds 42 (Q)SAR models from the DK-DB: For real-time predictions of user-defined structures and download of details in **QPRF**

Btw. the system has a structure curation module to ensure same format as used for modelling (neutralizing etc.)

CONCERTREACH
CONCERTING EXPERIMENTAL DATA AND IN SILICO MODELS FOR REACH
LIFE17 GIE/IT/000461

Novelty generated by DTU in this project so far

- Generate and include **predictions from a number of VEGA models in the Danish DB**
  - To choose the most relevant and best models to supplement the information in our DB we have thoroughly considered all VEGA model documentation (QMRFs and scientific publications) according to OECD validation principles and are in dialogue with IRFMN about our outstanding questions. So far we have chosen and implemented the Ames consensus mutagenicity models (including the four underlying models) and a skin sensitisation model. We are in the nearest future - depending on the outcome of the dialogue with IRFMN - going to choose and implement other models.
- Training in the Danish (Q)SAR Database
- Contributed information on our non-confidential QSAR training sets to IRFMN
- Contribute to scientific conference / meetings dissemination
- Organize this **workshop** for EU authorities
- Planning together with knoell **webinars / e-meetings**
- **Coordinate** communication and dissemination task

CONCERTREACH
CONCERTING EXPERIMENTAL DATA AND IN SILICO MODELS FOR REACH

LIFE17 GIE/IT/000461

Thank you

CONCERTREACH
CONCERTING EXPERIMENTAL DATA AND IN SILICO MODELS FOR REACH
LIFE17 GIE/IT/000461

Recent updates of the database and models websites

- The database and models websites are **continuously being expanded** regarding included models / information and updated regarding the user interface

- E.g. in 2020, we updated the systems with:
  - 8 additional models (1 CERAPP, 3 CoMPARA and 4 CAR)
  - Cumulated REACH registration minimum tonnages
  - EU CLP harmonized classifications
  - DK advisory classifications
  - Interface with lists of model statistics / QMRF links, reports, papers, presentations etc.
  - New advanced search functionality (min/max combinations of previous searches)

- In the **EU LIFE Concert REACH** project we've included **VEGA** predictions and more will follow

- Models under development by us in the **EU H2020 projects FREIA and ATHENA** as well as for **skin sensitization** in project for DK-EPA