

Implementations of VEGA predictions in the Danish (Q)SAR Database

Eva B. Wedebye, Nikolai G. Nikolov, Henrik Tyle, Ana C.V.E. Nissen

EU LIFE CONCEERT REACH Final Workshop 19 June 2023



DTU Food
National Food Institute

Outline

- DTU role in the EU LIFE Concert REACH project
- VEGA models implemented in the Danish (Q)SAR Database

DTU contributions in the LIFE Concert REACH project

- Generate and include **predictions from VEGA models in the Danish Database**
- Contribute to the **CONCERT Gateway**
- **Training** in the Danish (Q)SAR Database
- Publish non-confidential QSAR **training sets** to in the Danish (Q)SAR Database
- Contribute to **dissemination** (webinars, scientific conferences, meetings)
 - Incl. organizing a **workshop** for EU authorities (Oct 2021)

Chosen VEGA models for implementation in the Danish (Q)SAR Database

- Water solubility model (IRFMN)
- Air Half-Life (IRFMN/CORAL)
- Ready Biodegradability model (IRFMN)
- BCF model (CAESAR)
- Daphnia Magna Chronic (NOEC) toxicity model (IRFMN)
- Daphnia Magna Acute (EC50) Toxicity Model (IRFMN)
- Algae Acute (EC50) Toxicity model (IRFMN)
- Sludge (EC50) Toxicity Model (ProtoQSAR/Combase)
- Sludge Classification Toxicity Model (ProtoQSAR/Combase)
- MOA Toxicity classification (EPA T.E.S.T)
- Developmental/Reproductive Toxicity library (PG)
- Hepatotoxicity Model (IRFMN)
- In vitro Micronucleus activity (IRFMN)
- Mutagenicity (Ames test) Consensus
- Mutagenicity (Ames test) model (SarPy/IRFMN) Model
- Mutagenicity (Ames test) Model (ISS)
- Mutagenicity (Ames test) Model (KNN/Read-Across)
- Mutagenicity (Ames test) Model (CEASAR)

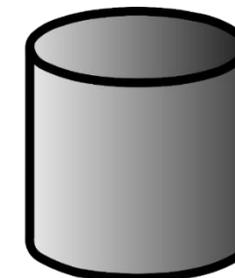
Predictions from VEGA models in the Danish Database

- To choose the most relevant and best models to **supplement the information** in the Danish (Q)SAR Database we have thoroughly considered **all VEGA models' documentation** (QMRFs and scientific publications) according to the OECD validation principles and in dialogue with IRFMN about questions
- Predictions are implemented following the general concept / terminology of the database, i.e. abbreviated predictions **POS/NEG** for binary endpoints, **AD IN/OUT**, however we added indication of VEGA calls good (IN) / moderate (OUT) / low (OUT) for transparency reasons
- All VEGA predictions are **searchable for predictions in domain**, and can thereby in very fast manner e.g. be crossed with REACH-registered substances etc. and other searches in making combined / complex searches
- **Launched today** in the Danish (Q)SAR Database

Danish (Q)SAR Database

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

- Developed and maintained by our DTU QSAR team
- Pre-generated predictions for **>650,000 mono-constituent organic substances**, including 13,406 REACH-registered substances
- **>200 QSAR DTU/commercial/free models** used
 - phys.chem / ADME / HH / ENV
- Documentation of models in (Q)SAR Model Reporting Format, **QMRF**
- **Free, easy-to-use, fast and advanced searches**
- Integrated with the free **OECD QSAR Toolbox**
- Danish (Q)SAR **Models** for detailed predictions in (Q)SAR Prediction Reporting Format, **QPRF**
- Both sites are **continuously being updated** with new models / substances



Database

**Statistics since release
November 2015**

>10k unique IPs made
>200k 'real' searches

Danish (Q)SAR Database

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

- Developed and maintained by our DTU QSAR team
- Pre-generated predictions for **>650,000 mono-constituent organic substances**

• **>200 C**

– phy

• Docume

• **Free, e**

• Integra

• Danish
Format

• Both sit

Environmental Toxicology and Pharmacology 98 (2023) 104069

Contents lists available at [ScienceDirect](#)

 **Environmental Toxicology and Pharmacology**

journal homepage: www.elsevier.com/locate/etap

A method for *in vitro* data and structure curation to optimize for QSAR modelling of minimum absolute potency levels and a comparative use case

Nikolai G. Nikolov¹, Ana C.V.E. Nissen, Eva B. Wedebye^{*,1}

- Transparent principles and reporting of data curation
- Transparency also for validation for external predictivity (e.g. no automatic exclusion of outliers, not using validation to choose model etc.)

(Q)SAR software/models

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



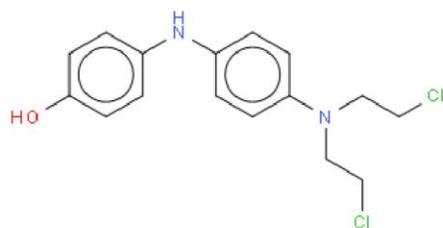
QSAR TOOLBOX

VEGAHUB

Validation and documentation

- QSAR Model Reporting Formats (QMRFs) for ~140 + 18 VEGA models in the Danish (Q)SAR Database, now also **linked in the EU LIFE CONCERT REACH Gateway**

<p>Schmieske QSAR model for the Bacterial Reverse Mutation Assay (Ames test)</p> <p>1. QSAR identifier</p> <p>1.1 QSAR identifier (short)</p> <p>Schmieske QSAR model for the Bacterial Reverse Mutation Assay (Ames test)</p> <p>1.2 Other related models</p> <p>Ames test QSAR model for the Bacterial Reverse Mutation Assay (Ames test)</p> <p>1.3 Software using the model</p> <p>QSAR version 3.1.00</p>	<p>2. General information</p> <p>2.1 Date of QMRF</p> <p>January 2015</p> <p>2.2 QMRF author(s) and contact information</p> <p>QSAR Group at DTU Food, Danish Institute for Food and Veterinary Research, Copenhagen, Denmark</p> <p>2.3 Name of QSAR update(s)</p> <p>2.4 QSAR update(s)</p> <p>2.5 Model developer(s) and contact information</p> <p>2.6 Software name and version for applicability domain</p>	<p>3. Defining the endpoint</p> <p>3.1 Species</p> <p>3.2 Endpoint</p> <p>3.3 Test method</p>	<p>4. Defining the algorithm</p> <p>4.1 Type of model</p> <p>4.2 Description of the model</p> <p>4.3 Type of model</p>	<p>5. Defining Applicability Domain</p> <p>5.1 Description of the applicability domain of the model</p> <p>5.2 Software used for applicability domain</p> <p>5.3 Method used to assess the applicability domain</p> <p>5.4 Software name and version for applicability domain</p>	<p>6. Internal validation</p> <p>6.1 Availability of the training set</p> <p>6.2 Available information for the training set</p> <p>6.3 Data for each descriptor variable for the training set</p> <p>6.4 Data for the dependent variable for the training set</p> <p>6.5 Other information about the training set</p> <p>6.6 Pre-processing of data before modeling</p> <p>6.7 Statistics for goodness-of-fit</p>	<p>7. External validation</p> <p>7.1 Availability of the external validation set</p> <p>7.2 Available information for the external validation set</p> <p>7.3 Data for each descriptor variable for the external validation set</p> <p>7.4 Other information about the training set</p> <p>7.5 Experimental design of test set</p> <p>7.6 Predictability – Assessment of the external validation set</p> <p>7.7 Comments on the external validation set</p>	<p>8. Mechanistic interpretation</p> <p>8.1 Mechanistic basis of the model</p> <p>8.2 A priori or posteriori mechanistic interpretation</p> <p>8.3 Other information about the mechanistic interpretation</p>	<p>9. Miscellaneous information</p> <p>9.1 Comments</p> <p>9.2 Bibliography</p> <p>9.3 Supporting information</p>
--	---	---	--	---	---	--	---	---

(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

Registry Number	63979-55-5	PubChem CID	
REACH EC Number (pre-registration, by 2013)		REACH EC Number (registration, 2019 or 2022)	
REACH registration (2022)		REACH registration cumulated minimum annual tonnage (2022)	
EU CLP Harmonized Classification*		DK-EPA / DTU QSAR-based CLP Advisory Classification	
EU Biocide active substances		EU Pesticide active substances	
EU EFSA Botanical substances		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)		Vapour Pressure Experimental (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

	Exp	Prediction	Domain
Water solubility v1.0.1 (mg/L)		4.45	mod_OUT
VEGA 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

Air Half-Life

	Exp	Prediction	Domain
Air Half-Life (CORAL) v1.0.1 (hr)		0.3282	low_OUT

VEGA 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 cm ³ /molecule-sec and OZ: E-17 cm ³ /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_IN	POS_IN	POS_IN	NEG_OUT

DTU-developed models

POS=Not Ready

Not Ready Biodegradability v1.0.10	POS_good_IN
VEGA model	
POS=Not Ready	

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

	Exp	Prediction	Domain
Log BCF (CAESAR) v2.1.15 (L/kg)		2.56	mod_OUT

VEGA model

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

	Exp	Prediction	Domain
Daphnia magna 48h EC50 v1.0.1 (mg/L)		0.0489	mod_OUT
Daphnia magna 21d NOEC v1.0.1 (mg/L)		0.309	low_OUT

Algae Acute 72h ErC50 v1.0.1 (mg/L)	0.362	low_OUT
Sludge Classification, 3h EC50 < 100 mg/L v1.0.1*	NEG_low_OUT	See prediction
Sludge 3h EC50 v1.0.1 (mg/L)	33.45	low_OUT

VEGA models

* The quantitative model should only be applied when the Sludge classification model is POS_IN

MOA Toxicity classification (EPA T.E.S.T) v1.0.2	
- predicted Mode of action	Narcosis
- Domain	good_IN
- Exp, if part of training set	

VEGA model

Qualitative estimation of MOA for Fathead minnow 96h

	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/cm ² /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.

Effects 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.

	Exp	Prediction_Domain
Hepatotoxicity v1.0.1		POS_mod_OUT

VEGA model

Profiler-type of predictions to be used as supporting information together with relevant QSAR predictions.

Irritation and Sensitization

	Exp	Battery	CASE Ultra	Leadscope	SciQSAR
Severe Skin Irritation in Rabbit		POS_IN	POS_IN	POS_IN	NEG_IN
Skin sensitisation GHS/CLP at least Cat. 1, LLNA-based (open data only)				POS_IN	
Skin sensitisation GHS/CLP at least Cat. 1, LLNA-based (open data and REACH-registrations)	N/A			POS_IN	
Skin sensitisation GHS/CLP at least Cat. 1, LLNA-based, only negative predictions (open data only)				N/A	
Skin sensitisation GHS/CLP Cat. 1A, LLNA-based (open data only)				POS_IN	
Skin sensitisation GHS/CLP Cat. 1A, LLNA-based (open data and REACH-registrations)	N/A			POS_IN	
Skin sensitisation GHS/CLP Cat. 1A, LLNA-based, only positive predictions (open data and REACH-registrations)	N/A			POS_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

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 <i>in vitro</i>)		INC_OUT	NEG_IN	INC_OUT	POS_IN
Estrogen Receptor α Binding, Balanced Training Set (Human <i>in vitro</i>)		POS_IN	POS_IN	POS_IN	NEG_IN
Estrogen Receptor α Activation (Human <i>in vitro</i>)		POS_OUT	INC_OUT	INC_OUT	POS_IN
Estrogen Receptor Activation, CERAPP data (<i>in vitro</i>)		N/A	N/A	NEG_IN	N/A
Androgen Receptor Inhibition (Human <i>in vitro</i>)		INC_OUT	INC_OUT	NEG_IN	POS_IN
Androgen Receptor Binding, CoMPARA data (<i>in vitro</i>)		N/A	N/A	POS_OUT	N/A
Androgen Receptor Inhibition, CoMPARA data (<i>in vitro</i>)		N/A	N/A	POS_IN	N/A

	Exp	Battery	CASE Ultra	Leadscope	SciQSAR
Androgen Receptor Activation, CoMPARA data (<i>in vitro</i>)		N/A	N/A	NEG_IN	N/A
Thyropoxidase (TPO) inhibition QSAR1 (Rat <i>in vitro</i>)		N/A	N/A	POS_IN	N/A
Thyropoxidase (TPO) inhibition QSAR2 (Rat <i>in vitro</i>)		N/A	N/A	POS_IN	N/A
Sodium/iodide symporter (NIS), higher sensitivity		N/A	N/A	POS_IN	N/A
Sodium/iodide symporter (NIS), higher specificity		N/A	N/A	INC_OUT	N/A
Thyroid Receptor α Binding (Human <i>in vitro</i>)					
- mg/L			52024.02	201.7734	69.71142
- μ M			159955.8	620.3831	214.3384
- Positive for IC ₅₀ \leq 10 μ M					
- Positive for IC ₅₀ \leq 100 μ M					
- Domain	OUT	OUT	OUT	OUT	OUT
Thyroid Receptor β Binding (Human <i>in vitro</i>)					
- mg/L			10524.57	6.459723	628.4979
- μ M			32359.38	19.86141	1932.413
- Positive for IC ₅₀ \leq 10 μ M					
- Positive for IC ₅₀ \leq 100 μ M					
- Domain	OUT	OUT	OUT	OUT	OUT
Arylhydrocarbon (AhR) Activation – Rational final model (Human <i>in vitro</i>)		N/A	N/A	NEG_IN	N/A
Arylhydrocarbon (AhR) Activation – Random final model (Human <i>in vitro</i>)		N/A	N/A	NEG_OUT	N/A
Pregnane X Receptor (PXR) Binding (Human <i>in vitro</i>)	N/A	INC_OUT	POS_OUT	POS_OUT	INC_OUT
Pregnane X Receptor (PXR) Binding (Human <i>in vitro</i>) NEW		N/A	N/A	NEG_IN	N/A
Pregnane X Receptor (PXR) Activation (Human <i>in vitro</i>)		N/A	N/A	INC_OUT	N/A
Pregnane X Receptor (PXR) Activation (Rat <i>in vitro</i>)		N/A	N/A	NEG_IN	N/A
CYP3A4 Induction (Human <i>in vitro</i>)		N/A	N/A	NEG_OUT	N/A
Constitutive Androstane Receptor (CAR) Activation at max. 20 μ M (<i>in vitro</i>)		N/A	N/A	POS_IN	N/A
Constitutive Androstane Receptor (CAR) Activation at max. 50 μ M (<i>in vitro</i>)		N/A	N/A	POS_IN	N/A
Constitutive Androstane Receptor (CAR) Inhibition at max. 20 μ M (<i>in vitro</i>)		N/A	N/A	NEG_IN	N/A
Constitutive Androstane Receptor		N/A	N/A	POS_IN	N/A

	Exp	Battery	CASE Ultra	Leadscope	SciQSAR
(CAR) Inhibition at max. 50 μ M (<i>in vitro</i>)					
DTU-developed models					
Estrogen Receptor Binding, alerts in:					
- parent only		Strong binder, OH group			
- metabolites from <i>in vivo</i> Rat metabolism simulator only		Strong binder, NH ₂ group; Strong binder, OH group; Moderate binder, NH ₂ 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 <i>in vivo</i> 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				

	Exp	Prediction_Domain
Developmental/Reproductive Toxicity library (PG) v1.1.2		NEG_low_OUT
VEGA model		
Profiler-type of predictions to be used as supporting information together with relevant QSAR predictions		

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	
Profilers 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 <i>S. typhimurium</i> (<i>in vitro</i>)		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 \geq 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.

	Consensus	Mut. / Non-mut. Scores	Used models
Ames test Consensus	POS	0.675 / 0	4

VEGA model

Mutagenicity (Ames) consensus model version 1.0.4 contained in VEGA version 1.4.3 with calculation core version 1.3.14

	Exp	Prediction
Ames test (ISS) v1.0.3		POS_good_IN
Ames test (CAESAR) - v2.1.13		POS_mod_OUT
Ames test (SarPy) v1.0.8		POS_mod_OUT
Ames test (KNN/Read-Across) v1.0.1		POS_mod_OUT

VEGA models

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

	Exp	Prediction_Domain
Micronucleus (VERMEER) v1.0.1		POS_mod_OUT
VEGA model		

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 <i>Drosophila m.</i>		POS_IN	POS_IN	POS_IN	POS_IN

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 from 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, VEGA 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\pLogD

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: <https://www.multicase.com/case-ultra>

Leadscope: <https://www.leadscope.com/>

VEGA: <https://vegahub.it>

ToxSuite: <https://www.acdlabs.com/>

Copyright notice, terms and conditions of use

Copyright notice, terms and conditions of use

Permission is granted to use information from the database as is. The database is an expert tool where the final assessment of properties is not dictated by the (Q)SAR estimates, but by the user's own scientific judgment. Aside from the fact that models are never perfect, the (Q)SAR field is under rapid development and models are regularly updated and improved. It is also impossible to provide the detailed information accompanying each individual prediction that is available to those who do not own licences to the software platforms. The structural information in the database stems from many sources and in some cases it may be wrong. The structures are also in some cases abbreviated in that possible anions and cations have been removed. This can have important toxicological significance (e.g. for Heavy Metal salts).

All access to the database should happen through the provided client-side software and without any use of automated workflow or scripting.

Reproduction of information from the database is permitted provided the source is acknowledged as follows:

Danish (Q)SAR Database, National Food Institute, Technical University of Denmark, <https://qsar.food.dtu.dk>.

The Technical University of Denmark (DTU) is not responsible for any errors or inaccuracies the database may contain and is not liable for any use that may be made of the information contained therein. DTU do not warrant, and hereby disclaim any warranties, with respect to the accuracy, adequacy or completeness of any information obtained from this database. Nor do we warrant that the site will operate in an uninterrupted or error-free manner or that the site and its components are free of viruses or other harmful components. Use of information obtained from or through this site is at your own risk.

As a user of this database, you agree to indemnify and hold DTU harmless from any claims, losses or damages, including legal fees, resulting from your use of this database, and to fully cooperate in DTU's defense against any such claims.

The user requests are processed by the server hosting the database which in the process stores information. Only authorized employees have authorized access to the server and reasonable measures are in place to protect the server from unauthorized access. DTU uses the stored user request information solely for error tracking and to collect anonymized statistics (number of users, number of searches, number of report downloads etc.), and we do not release any information at the level of individual searches. However, as any server/PC/network that the requests pass through may be compromised by unauthorized access, we cannot guarantee that the information submitted by users does not fall into the hands of third parties.

Danish (Q)SAR Database



Ministry of Environment and Food of Denmark
Environmental Protection Agency

DTU Food
National Food
Institute



New search

Search input field with buttons: ID, Structure and name

Search filters: PhysChem, Environment, ADME, Human health

Search logic: AND Intersect results, OR Unite results

Copyright notice, terms and conditions of use

Permission is granted to use information from the database as is. The database is an expert tool where the final assessment of properties is not dictated by the (Q)SAR estimates, but by the user's own scientific judgment. Aside from the fact that models are never perfect, the (Q)SAR field is under rapid development and models are regularly updated and improved. It is also impossible to provide the detailed information accompanying each individual prediction that is available to those who do not own licences to the software platforms. The structural information in the database stems from many sources and in some cases it may be wrong. The structures are also in some cases abbreviated in that possible anions and cations have been removed. This can have important toxicological significance (e.g. for Heavy Metal salts).

All access to the database should happen through the provided client-side software and without any use of automated workflow or scripting.

Reproduction of information from the database is permitted provided the source is acknowledged as follows:

Danish (Q)SAR Database, Research Group for Chemical Risk Assessment and GMO, National Food Institute, Technical University of Denmark, <http://qsar.food.dtu.dk>.

The Technical University of Denmark (DTU) is not responsible for any errors or inaccuracies the database may contain and is not liable for any use that may be made of the information contained therein. DTU do not warrant, and hereby disclaim any warranties, with respect to the accuracy, adequacy or completeness of any information obtained from this database. Nor do we warrant that the site will operate in an uninterrupted or error-free manner or that the site and its components are free of viruses or other harmful components. Use of information obtained from or through this site is at your own risk. As a user of this database, you agree to indemnify and hold DTU harmless from any claims, losses or damages, including legal fees, resulting from your use of this database, and to fully cooperate in DTU's defense against any such claims.

The user requests are processed by the server hosting the database which in the process stores information. Only

I agree

Danish (Q)SAR Database



Home Clear **Information** Contact QSAR2023

New search

ID

Structure and n

PhysChem

Environment

ADME

Human health

User manual

Model documentation

Brochure

Scientific articles

Reports

Presentations

Disclaimer



Danish (Q)SAR Database



Home Clear Information Contact QSAR2023

[Back to search](#)

Endpoint	N in training set	(Cross) validation result (%) ^a	QMRF	Training sets
Water solubility v1.0.1 (mg/L)	4014	Test set in AD: n=341; Q ² =0.91; RMSE=0.69	VEGA	
Air Half-Life (CORAL) v1.0.1 (hr)	76+77+74	Test set in AD: n=34; Q ² =0.89; RMSE=0.29	VEGA	
Not ready biodegradability (POS=Not Ready)	735	Sens=68.9, Spec=87.8, Conc=77.2	CASE Ultra	
		Sens=87.3, Spec=85.2, Conc=86.4	Leadscope	
		Sens=63.0, Spec=92.7, Conc=77.8	SciQSAR	
Ready Biodegradability model v1.0.10 (POS=Not Ready)	582	Test set in AD: n=71, Sens=100, Spec=87, BA=94	VEGA	
Log BCF (CAESAR) v2.1.15 (L/kg)	378	Test set in AD: n=31; Q ² =0.85; RMSE=0.52	VEGA	
Fathead minnow 96h LC50 (mg/L)	565	R ² =0.75, Q ² =0.73	Leadscope	
		R ² =0.74, Q ² =0.72	SciQSAR	
Daphnia magna 48h EC50 (mg/L)	626	R ² =0.67, Q ² =0.64	Leadscope	
		R ² =0.65, Q ² =0.63	SciQSAR	
Pseudokirchneriella s. 72h EC50 (mg/L)	531	R ² =0.74, Q ² =0.71	Leadscope	
		R ² =0.64, Q ² =0.60	SciQSAR	
Daphnia magna 48h EC50 v1.0.1 (mg/L)	312	Test set in AD: n=44; Q ² =0.66; RMSE = 0.60	VEGA	

- Training sets for DTU models launched later today

Danish (Q)SAR Database



Home Clear Information Contact QSAR2023

New search

Searches



Results

Sub

ID

Structure and name

PhysChem

Environment

- Acute and Repeated dose toxicity ▶
- Irritation and Sensitisation ▶
- Endocrine and Molecular Endpoints ▶
- Developmental Toxicity ▶
- Genotoxicity ▶**
- Carcinogenicity (in vivo) ▶

OR

Unite results

NOT

Complement results

MAX/MIN

- DNA Reactivity ▶
- Ames test ▶
- Other in vitro endpoints ▶**
- In vivo endpoints ▶

- Chromosome Aberrations in Chinese Hamster Ovary Cells
- Chromosome Aberrations in Chinese Hamster Lung Cells
- Mutations in Thymidine Kinase Locus in Mouse Lymphoma Cells
- Mutations in HGPRT Locus in Chinese Hamster Ovary Cells
- Unscheduled DNA Synthesis in Rat Hepatocytes
- Syrian Hamster Embryo Cell Transformation
- VEGA In vitro Micronucleus activity (IRFMN) v1.0.1**
- Profilers ▶

New search

ID

Structure and name

PhysChem

Environment

ADME

Human health

AND
Intersect results

OR
Unite results

NOT
Complement results

MAX/MIN

Searches



VEGA In vitro Micronucleus activity (IRFMN) v1.0.1

Search Model documentation

VEGA In vitro Micronucleus activity (IRFMN) v1.0.1

Select predictions or experimental results:

VEGA
Experimental (from training set)

and search for structures predicted:

Positive
in applicability domain

Negative
in applicability domain

Cancel

Danish (Q)SAR Database



Ministry of Environment and Food of Denmark
Environmental Protection Agency



Home Clear Information Contact QSAR2023

Danish (Q)SAR Models

New search

ID

Structure and name

PhysChem

Environment

ADME

Human health

AND

Intersect results

Searches



1. POS VEGA In vitro Micronucleus activity (IRFMN)...
2. REACH registered substances (by 2022)
3. 1. AND 2.

Results

41392

13406

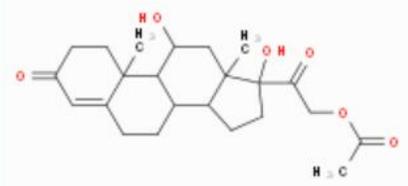
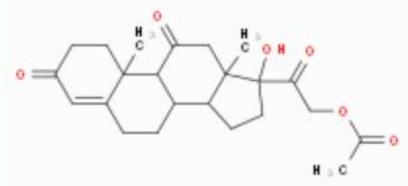
840

Substances

1. AND 2.: Page 1

Previous Next 1 2 3 84

Structures 1-10 of 840

Structure	Id	Similarity	+
	XX-XX-X		
	XX-XX-X		

Danish (Q)SAR Database



Home Clear Information Contact QSAR2023

New search

Searches < >

Results

Substances

ID

Structure and name

PhysChem

Environment

- Acute and Repeated dose toxicity ▶
- Irritation and Sensitisation ▶
- Endocrine and Molecular Endpoints ▶
- Developmental Toxicity ▶
- Genotoxicity ▶**
- Carcinogenicity (in vivo) ▶

OR

Unite results

NOT

Complement results

MAX/MIN

- DNA Reactivity ▶
- Ames test ▶**
- Other in vitro endpoints ▶
- In vivo endpoints ▶

- Bacterial Reverse Mutation Test (Ames test in *S. typhimurium* (in vitro))
- Direct Acting Ames Mutagens (without S9) - ONLY use for Ames POS_IN
- Base-Pair Ames Mutagens - ONLY use for Ames POS_IN
- Frameshift Ames Mutagens - ONLY use for Ames POS_IN
- Potent Ames Mutagens (Reversions ≥ 10 Times Controls) - ONLY use for Ames POS_IN
- Mutagenicity (Ames test), VEGA**
- Profilers ▶

Danish (Q)SAR Database



Home Clear Information Contact QSAR2023

New search

Searches < >

Results

ID
Structure and name

PhysChem
Environment
ADME
Human health

AND
Intersect results
OR
Unite results
NOT
Complement results
MAX/MIN

Mutagenicity (Ames test), VEGA

Search Model documentation

Mutagenicity (Ames test), VEGA

Select predictions or experimental results:

- VEGA Mutagenicity consensus model
- VEGA Mutagenicity model (CAESAR)
- Experimental (CAESAR model training set)
- VEGA Mutagenicity model (kNN)
- Experimental (kNN model training set)
- VEGA Mutagenicity model (IC)

and search for structures predicted:

Positive
in applicability domain

Negative
in applicability domain

Danish (Q)SAR Database



Ministry of Environment and Food of Denmark
Environmental Protection Agency

DTU Food
National Food
Institute



Home Clear Information Contact QSAR2023

Danish (Q)SAR Models

New search

ID

Structure and name

PhysChem

Environment

ADME

Human health

AND

Intersect results

Searches

1. POS Mutagenicity (Ames test), VEGA

Results

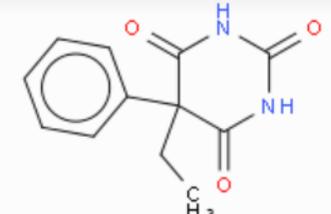
242735

Substances

POS Mutagenicity (Ames test), VEGA: Page 1

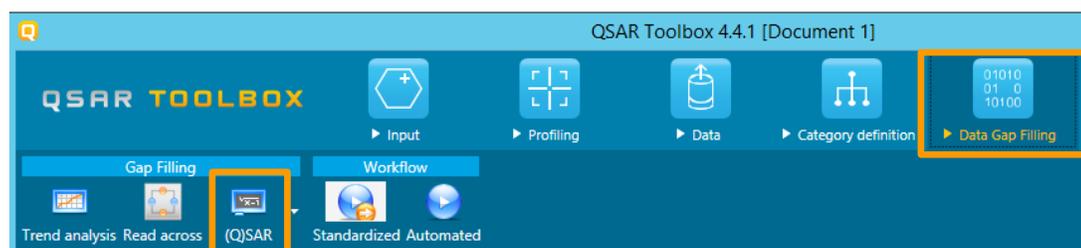
Previous Next 1 2 3 24274

Structures 1-10 of 242735

Structure	Id	Similarity	+
	XX-XX-X		↓
	XX-XX-X		↓

Integration with the OECD QSAR Toolbox in 2018

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



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** – 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).

Danish (Q)SAR Models

powered by Leadscope Predictive Data Miner



Ministry of Environment
and Food of Denmark



DTU Food
National Food
Institute



[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 50 (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.)

VEGA HUB

VEGA QSAR

Try out VEGA QSAR
and its features



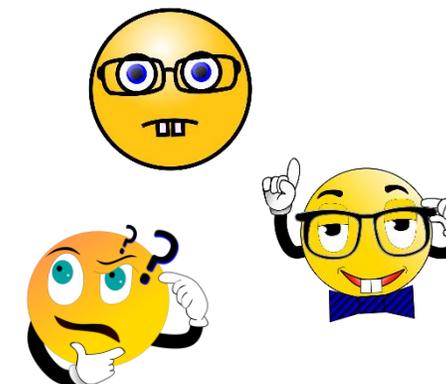
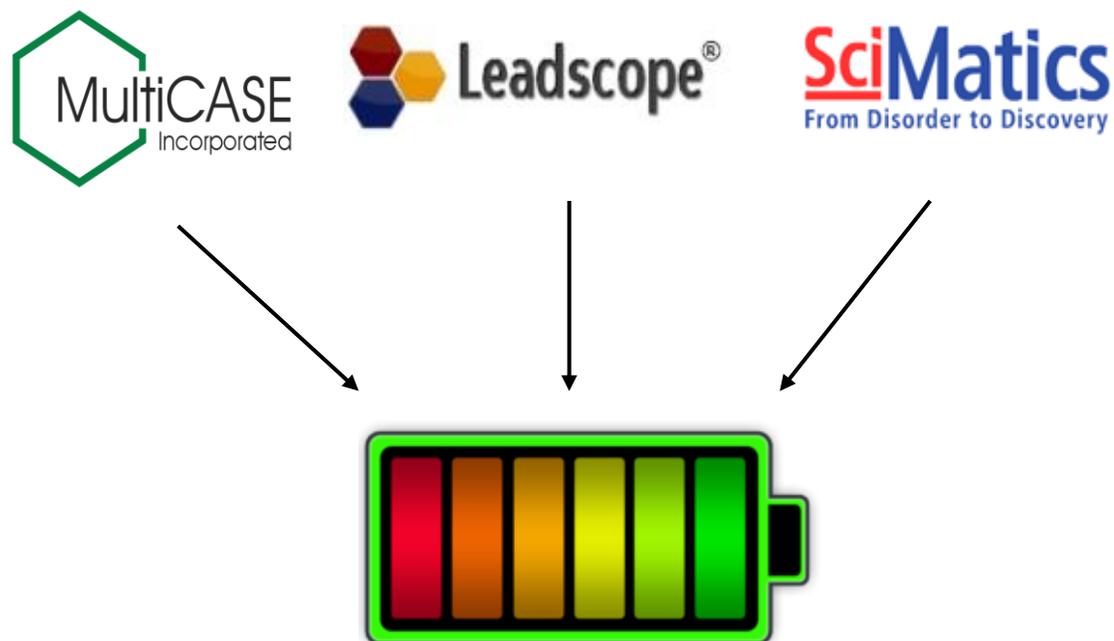
All the VEGA models are also available in a unique stand-alone application.

Thank you

Acknowledgements

EU LIFE CONCERT REACH for implementation of VEGA models
Danish EPA for 25y general support and many projects to make the Danish (Q)SAR DB
EU projects to expand the DB with new models
Nordic projects for development of the DB and WS for use
ECHA for support and use
EFSA for use for coming Botanicals database
Etc.

Development of battery approach for many training sets



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