CASE STUDY: using and integrating two QSAR models for assessing *in vitro* mutagenicity in bacteria

17/05/2023, LIFE CONCERT REACH Web-Seminars - (Q)SAR Models under REACH: Practical Examples

Rodolfo Gonella Diaza, Ph.D.
In silico expert
knoell Germany GmbH
qsar@knoell.com

CONCERTREACH
CONCERTING EXPERIMENTAL DATA AND IN SILICO MODELS FOR REACH

LIFE17 GIE/IT/000461

TABLE OF CONTENTS

Case study description

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

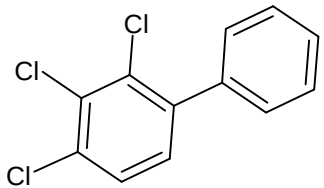
Aim: Prediction of *in vitro* gene mutation in bacteria with **two (Q)SAR models** and results **documentation in IUCLID**

Target molecule: 2,3,4-Trichlorobiphenyl

Knowledge-based model: VEGA – Mutagenicity (Ames test) model (ISS)

Statistical model: Danish (Q)SAR Database - Bacterial reverse mutation test (Ames test in S. typhimurium in vitro) – Battery model

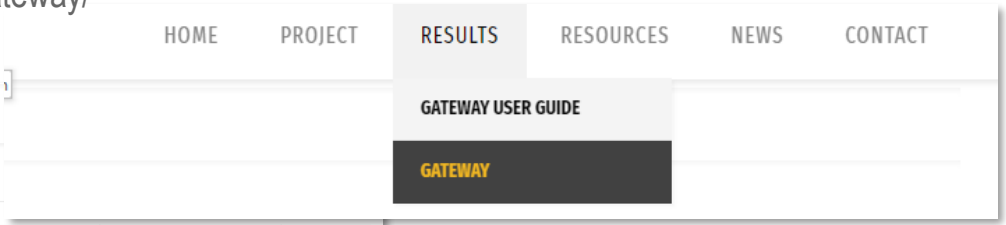
Select input method: SMILES notation - c1ccc(cc1)c2ccc(c(c2Cl)Cl)Cl



Models for *in vitro* gene mutation in bacteria in the CONCERT REACH gateway

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

https://www.life-concertreach.eu/results/results-gateway/



8.4.1. In vitro gene mutation study in bacteria

All | VEGA AND ToxRead | DANISH QSAR DATABASE | AMBIT | OCHEM

End Point	Model	Type	Dataset size	Training set size	Test set size	Cross-validation procedure	Platform	Remarks
TOX 7.6.1. Genetic toxicity in vitro. Ames test (OECD 471)	Mutagenicity (Ames test) CONSENSUS model	classification	0	0	0		VEGA	
TOX 7.6.1. Genetic toxicity in vitro. Ames test (OECD 471)	Mutagenicity (Ames test) model (CAESAR)	classification	4204	3367	837		VEGA	
TOX 7.6.1. Genetic toxicity in vitro. Ames test (OECD 471)	Mutagenicity (Ames test) model (KNN-Read-Across)	classification	5770	5770	0		VEGA	

Currently **22 Models** from VEGA and Danish (Q)SAR Database

Direct link to documentation in QSAR Model Reporting Format (**QMRF**)

Includes all information on model, for IUCLID entries

Models for *in vitro* gene mutation in bacteria

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

VEGA: four models for Ames test

- CAESAR - **Hybrid** model (statistical + knowledge-based)
- KNN-Read-Across - **read-across** model
- ISS - **knowledge-based** structural **alerts** (Benigni-Bossa rulebase)
- SarPy-IRFMN - **statistical** structural **alerts**

→ **CONSENSUS model**

Combining the 4 outcomes

Danish (Q)SAR Database: 15 statistical models and 2 knowledge-based alert profilers for Ames test

- Bacterial reverse mutation test (Ames test in S. typhimurium in vitro)

To be considered only if the Bacterial reverse mutation test model gives a **positive in domain outcome**

- Direct acting Ames mutagens (without S9)
- Base pair Ames mutagens
- Frame shift Ames mutagens
- Potent Ames mutagens, reversions ≥ 10 times controls

→ 3 models for each endpoint/mechanism +

Battery model

Combining the 3 outcomes

2 **knowledge-based profilers** from OECD QSAR Toolbox

DNA alerts for AMES by OASIS, alerts in parent only (OECD QSAR Toolbox v.4.2 Profiler)

In vitro mutagenicity (Ames test) alerts by ISS, alerts in parent only (OECD QSAR Toolbox v.4.2 Profiler)

Models for *in vitro* gene mutation in bacteria

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

How to select the appropriate model(s) for my substance?

- ***A priori*** selection is generally **not possible**
- However, **experience in using the models** might suggest which could give more reliable results for certain type of substances (e.g., industrial chemicals, active substances, etc.)
- Information on **compliance** of the target molecule **with the applicability domain of the model**
- **Comparison with similar molecules** with available experimental results
- **It is generally required to use multiple and different models for evaluating the same endpoint ***

Expert analysis of the results and supporting information is needed

Both VEGA and Danish (Q)SAR database provide information on applicability domain compliance and similar molecules can be extracted and analyzed

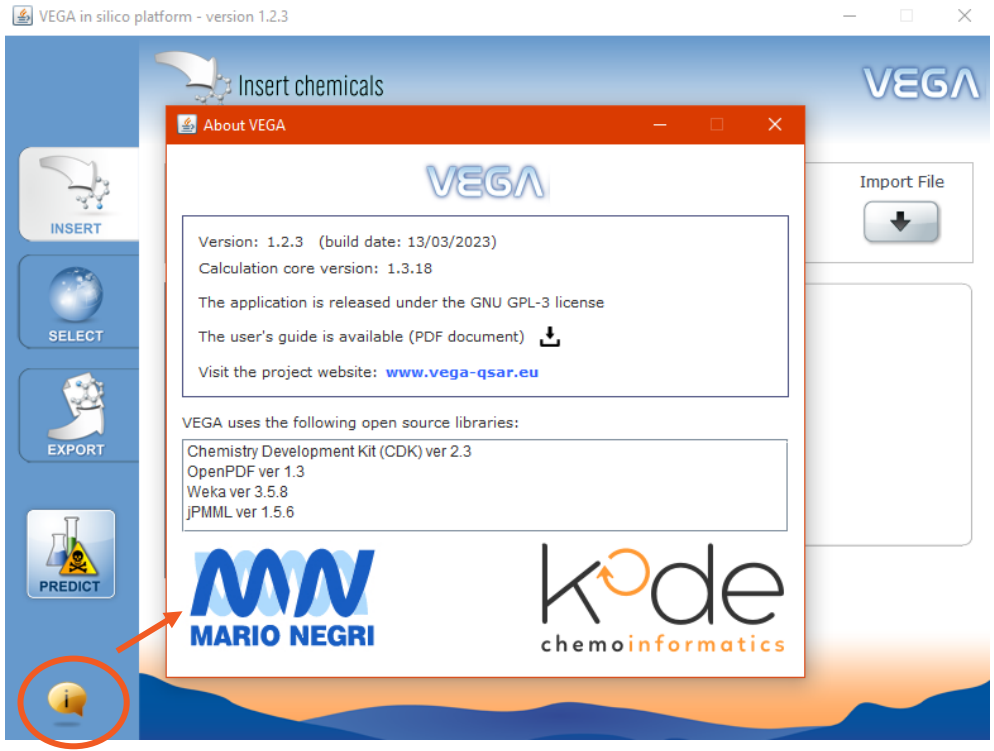
* additional info in appendix of this presentation

CONCERTREACH
CONCERTING EXPERIMENTAL DATA AND IN SILICO MODELS FOR REACH

LIFE17 GIE/IT/000461

TABLE OF CONTENTS

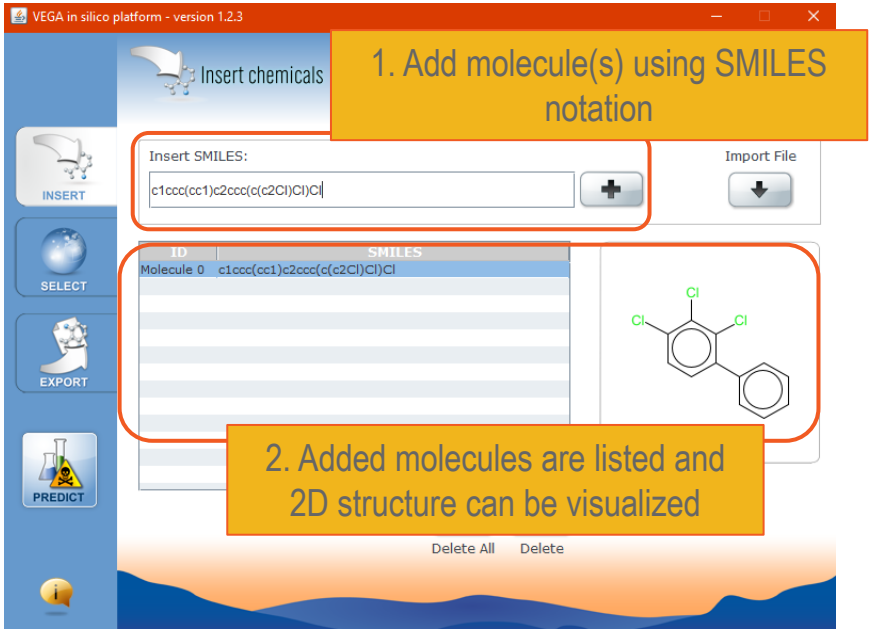
VEGA: introduction

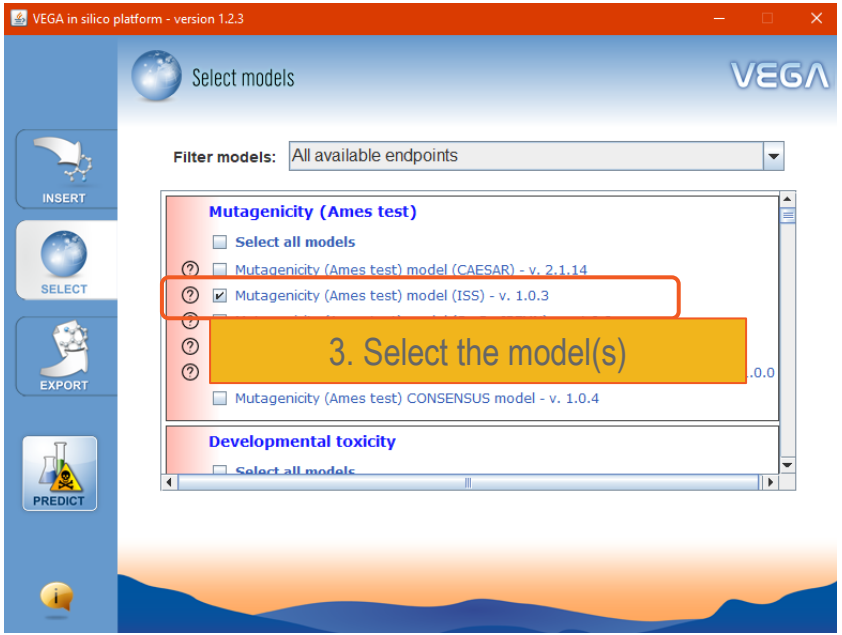


VEGA: Virtual models for Evaluating the properties of chemicals within a Global Architecture

- Developed mainly by Mario Negri Institute (Milan) and Kode s.r.l. (Pisa)
- **Free platform** developed based on contributions from EU projects
- Includes **110 statistical and knowledge-based (Q)SAR models** for the prediction of (eco)toxicity, environmental fate and physico-chemical properties of chemicals.

VEGA: running predictions

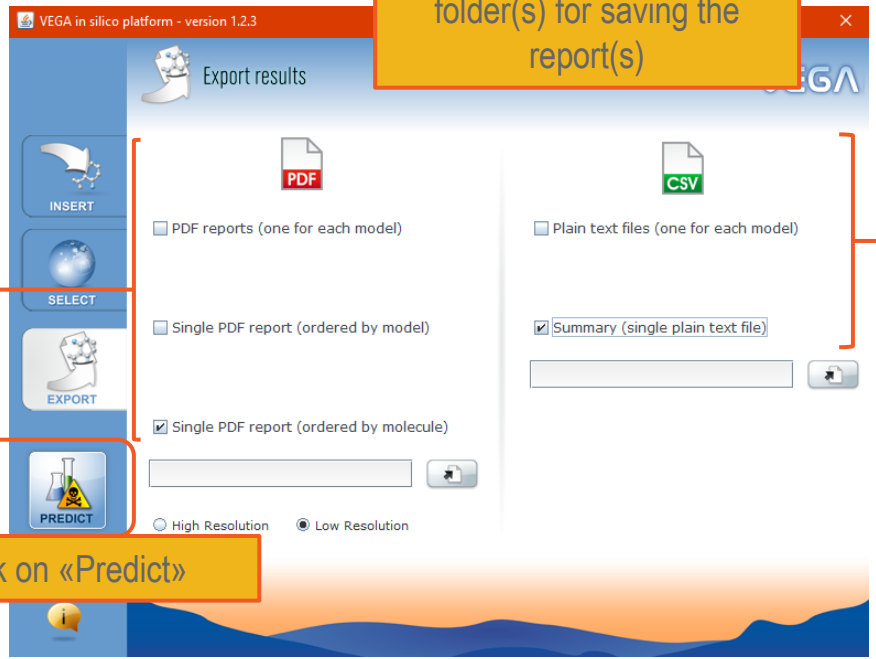




VEGA: running predictions

Full PDF reports:

- prediction(s) results
- applicability domain
- experimental data of the target (if any)
- most similar substances
- other supporting info (if any)



Simplified text reports

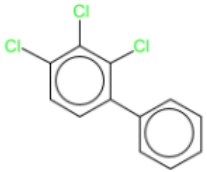
(useful for excel import)

VEGA: analysis of the results

Knowledge-based model: Mutagenicity (Ames test) model (ISS)

1. Prediction Summary

Prediction for compound Molecule 0 -



Prediction: Reliability:

Prediction is NON-Mutagenic, but the result shows some critical aspects, which require to be checked:
- some similar molecules found in the training set have experimental values that disagree with the predicted value

Compound: Molecule 0
Compound SMILES: c1ccc(cc1)c2ccc(c(c2Cl)Cl)Cl
Experimental value: -
Predicted Mutagen activity: NON-Mutagenic
Structural Alerts: -
Reliability: The predicted compound could be out of the Applicability Domain of the model
Remarks:
none

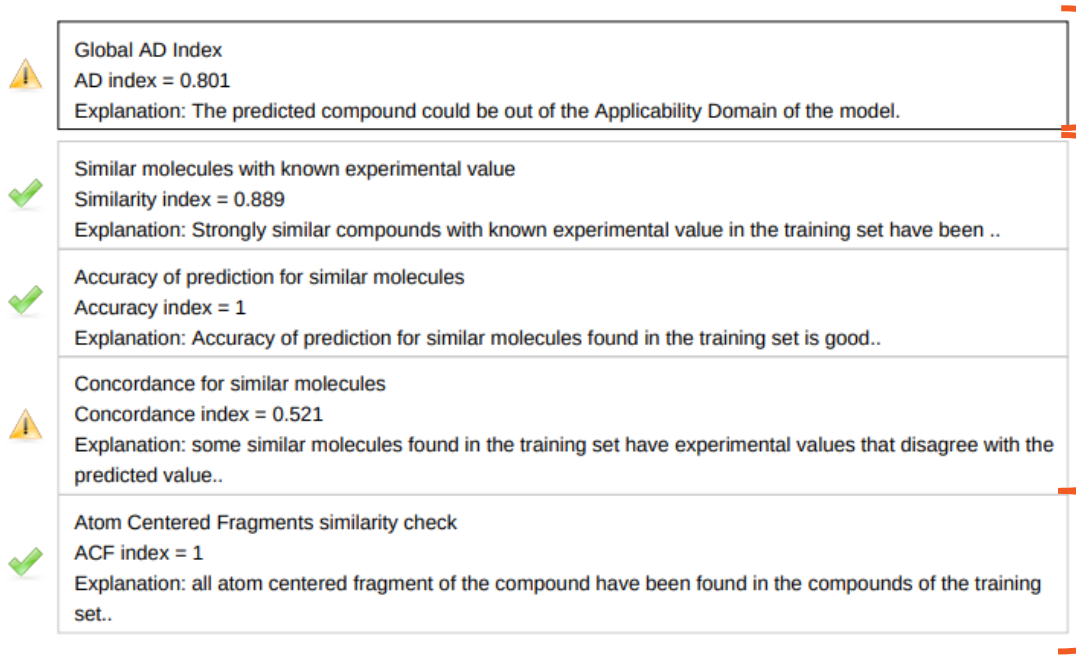
- Colour code for prediction (e.g., red = toxic, green = non-toxic)
- Stars for reliability from 1 (low) to 3 (high)
- Summary of the evaluation

- More details about the prediction, including **identified alerts (if any)**

The reliability of the prediction **is based on** an automated check of the molecule **compliance with the applicability domain of the model**.

VEGA: analysis of the results

Applicability Domain Index (ADI) ranges **from 0 (not in AD) to 1 (in AD)**
The ADI is calculated based on other indices, **each one taking into account a particular issue** of the applicability domain (AD)



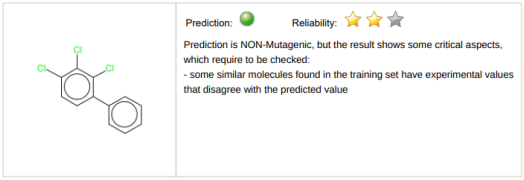
In AD if AD index ≥ 0.9

Calculated **based on the 2 most similar molecules** (not on all 6 reported similar molecules)

Considers the **whole training set** of the model

Number of considered similar molecules, number and type of indexes and thresholds are **model dependent**
Info in QMRF

VEGA: Example of critical evaluation of the automated AD / reliability evaluation



Compound: Molecule 0
Compound SMILES: c1ccc(cc1)c2ccc(c(c2Cl)Cl)Cl
Experimental value: -
Predicted Mutagen activity: NON-Mutagenic
Structural Alerts: -
Reliability: The predicted compound could be out of the Applicability Domain of the model
Remarks:
none

Global AD Index
AD index = 0.801
Explanation: The predicted compound could be out of the Applicability Domain of the model.

Similar molecules with known experimental value
Similarity index = 0.889
Explanation: Strongly similar compounds with known experimental value in the training set have been ..

Accuracy of prediction for similar molecules
Accuracy index = 1
Explanation: Accuracy of prediction for similar molecules found in the training set is good..

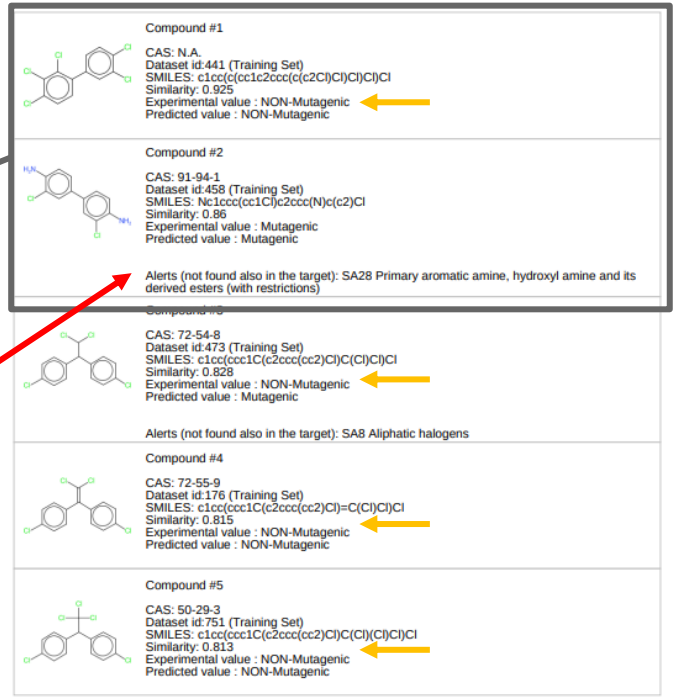
Concordance for similar molecules
Concordance index = 0.521
Explanation: some similar molecules found in the training set have experimental values that disagree with the predicted value..

Atom Centered Fragments similarity check
ACF index = 1
Explanation: all atom centered fragment of the compound have been found in the compounds of the training set..

Two most similar molecules considered

Affected by one molecule, **with a different alerts profile**

3.1 Applicability Domain:
Similar Compounds, with Predicted and Experimental Values



A **higher reliability could be assigned** to the negative prediction, also considering that **all other similar molecules** (mostly with the same “no alerts” profile) **are experimentally negative**

VEGA: take home message

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

- **Full documentation** of all models is available, in **QMRF format**
- Supporting information (**AD compliance, similar molecules**) is provided, **allowing expert evaluation**

Relevant for REACH dossier preparation in IUCLID

- AD compliance is **affected by identified similar molecules**
- Current **similarity evaluation does not consider important parameters** (e.g. different alerts, leading to different mechanism)
- **Automated AD compliance check** is not perfect and **requires user critical check**
  - ➢ This affects other tools as well, including commercial ones

A novel tool called VERA has been developed, aiming also at improving similarity evaluation and AD compliance check
(Next presentation, today)

CONCERTREACH
CONCERTING EXPERIMENTAL DATA AND IN SILICO MODELS FOR REACH

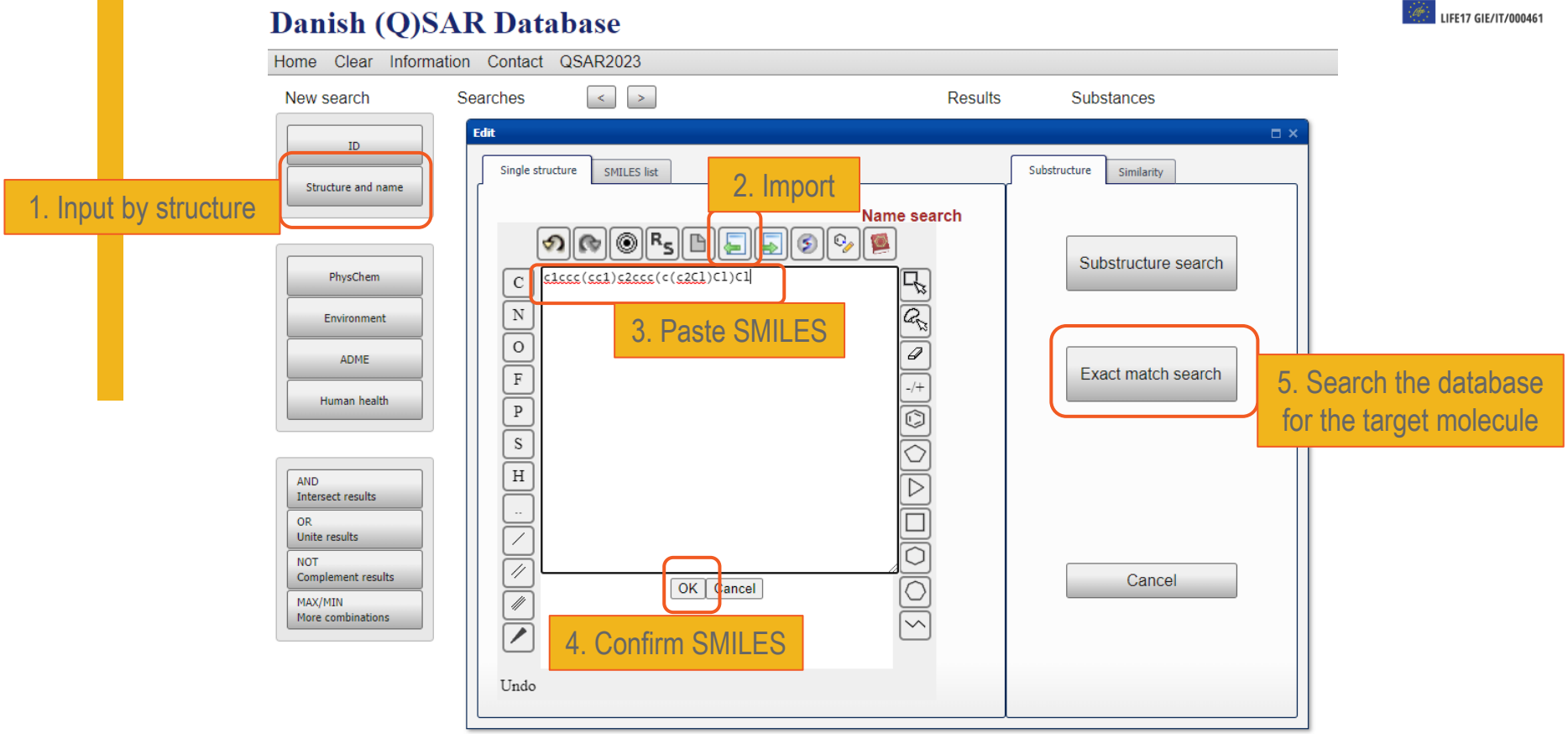
LIFE17 GIE/IT/000461

TABLE OF CONTENTS

Danish (Q)SAR Database: introduction

- Developed by the Danish Technical University
- **Freely accessible online database of (Q)SAR predictions** for a little over **650.000** chemical organic mono-constituent structures
- For each single contained substance, the database generates a report upon user request, which contains **more than 200 (Q)SAR results**
- Offers **advanced searching tools** for identifying potential **similar molecules for supporting (Q)SAR predictions** or candidate analogues for read-across

Danish (Q)SAR Database: gathering (Q)SAR results



Danish (Q)SAR Database: gathering (Q)SAR results

CONCERTREACH
CONCERTING EXPERIMENTAL DATA AND IN SILICO MODELS FOR REACH

LIFE17 GIE/IT/000461

New search

ID

Structure and name

PhysChem

Environment

ADME

Human health

AND
Intersect results

OR
Unite results

NOT
Complement results

MAX/MIN
More combinations

Searches

Results

Substances

1 Exact match: 1

6. Summary of the search performed

Exact match: : Page 1

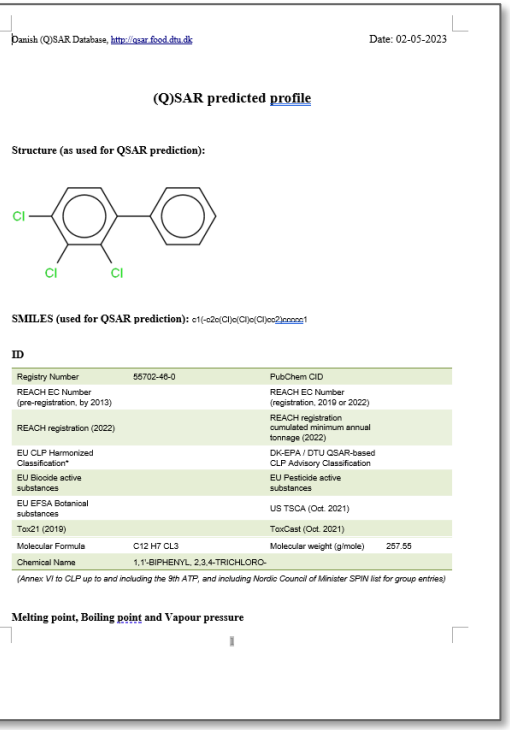
Previous Next 1

Structures 1-1 of 1

Structure | Id | Similarity | +

55702-XX-X

7. For each identified molecule, the (Q)SAR report can be downloaded in .rtf format



Danish (Q)SAR Database, http://qsar.food.dtu.dk Date: 02-05-2023

(Q)SAR predicted profile

Structure (as used for QSAR prediction):

SMILES (used for QSAR prediction): c1(-c2c(Cl)c(Cl)c(Cl)cc2)ccccc1

ID

Registry Number	55702-46-0	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	C12 H7 CL3	Molecular weight (g/mole)	257.55
Chemical Name	1,1'-BIPHENYL, 2,3,4-TRICHLORO-		

(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

Danish (Q)SAR Database: results for *in vitro* gene mutation in bacteria

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

3 models + battery (consensus) for Ames test and for the four further endpoints to be considered only if the outcome for Ames is Positive and in domain (POS_IN)

The target molecule was evaluated as **compliant with AD of all Ames models**, which generated **consistent negative predictions**. The other four models should not be considered.

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

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

Within LIFE CONCERT REACH, results from the four VEGA models and the Consensus model have been integrated

	VEGA	Mut. / Non-mut. scores	Used models
Mutagenicity consensus	NEG	0.23 / 0.25	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
NEG_Mod	NEG_Low	NEG_Low	POS_Good

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.

Structural alerts identified by two endpoint-specific profilers present in the OECD QSAR Toolbox

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	No alert found

OECD QSAR Toolbox v.4.2 profilers

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

Danish (Q)SAR Database: identification of similar molecules

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

- **Stepwise approach ***

- Danish (Q)SAR Database **can be searched** for molecules, based on available **experimental data, (Q)SAR predictions, structural alerts**, etc., for the endpoint of interest

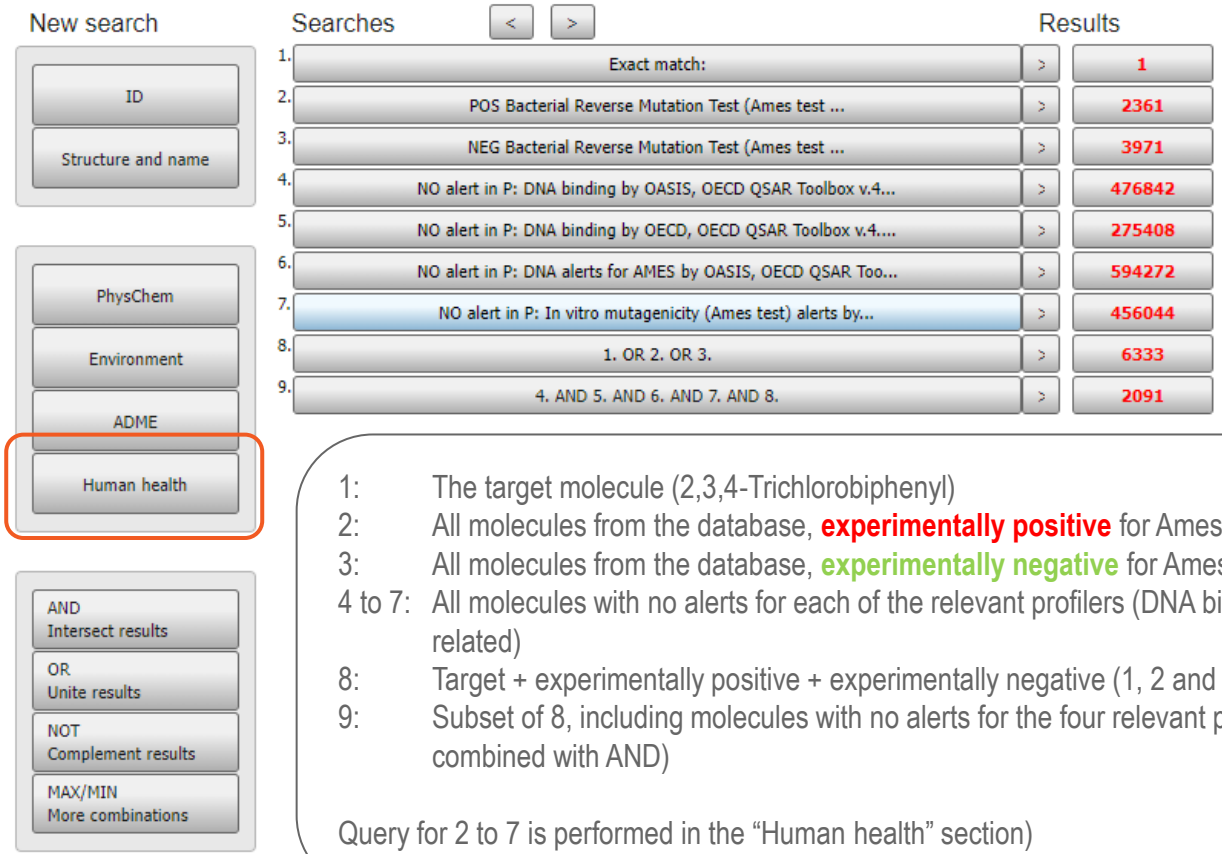
Our case:
- Search for experimentally **positive and negative** molecules for Ames;
- Target molecule has no Alerts for DNA binding or in vitro gene mutation in bacteria, **similar molecules can be selected with the same** "no alerts" **profile**.

- For each query, a **list of molecules is retrieved**

- The lists **can be merged**, using logical operator such as **AND** or **OR**

* Example of stepwise approach provided in appendix of this presentation

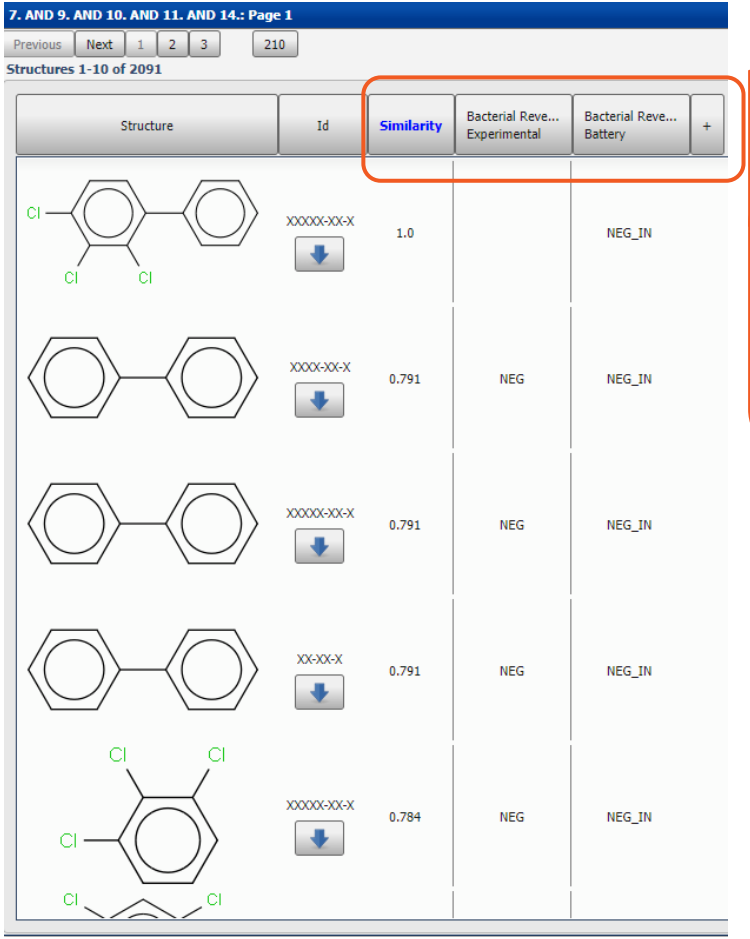
Danish (Q)SAR Database: identification of similar molecules



1: The target molecule (2,3,4-Trichlorobiphenyl)
2: All molecules from the database, **experimentally positive** for Ames test
3: All molecules from the database, **experimentally negative** for Ames test
4 to 7: All molecules with no alerts for each of the relevant profilers (DNA binding-related, Ames test-related)
8: Target + experimentally positive + experimentally negative (1, 2 and 3 combined with OR)
9: Subset of 8, including molecules with no alerts for the four relevant profilers (8, 4, 5, 6 and 7, combined with AND)

Query for 2 to 7 is performed in the “Human health” section)

Danish (Q)SAR Database: identification of similar molecules



Click to calculate similarity with the target molecule and automatically sorting by highest similarity
Use "+" to add columns to the overview

Columns can be added to show:
- Experimental data for the target endpoint
- Predictions for the target endpoint
- Other information useful for evaluating similar molecules (e.g. structural alerts from OECD QSAR Toolbox profilers)

The generated overview **can be used** to prepare a statement **for supporting the reliability** of the (Q)SAR prediction for the target molecule.

Danish (Q)SAR Database: take home message

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

- **Full documentation** of the models is available, in **QMRF format**
- Supporting information (**AD compliance, similar molecules**) is either provided or can be retrieved, **allowing expert evaluation**
- Identification of similar molecules is not automated; however, **more information can be taken into account** compared to other tools (e.g. alerts profile)
- The generated reports do not include full applicability domain compliance evaluation; however, the QMRF includes the definition of the applicability domain.

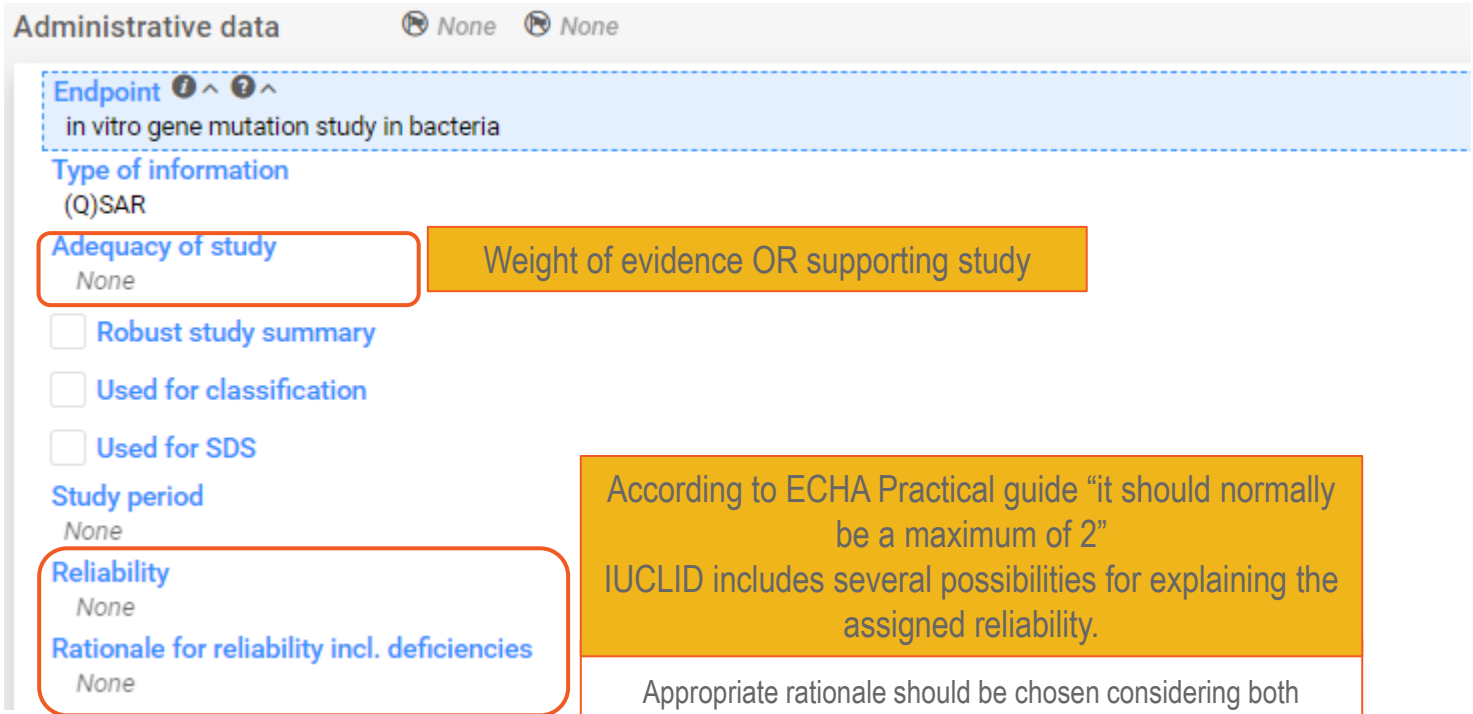
CONCERTREACH
CONCERTING EXPERIMENTAL DATA AND IN SILICO MODELS FOR REACH

LIFE17 GIE/IT/000461

TABLE OF CONTENTS

(Q)SAR results in IUCLID

VEGA outcome reported according to ECHA Practical guide “How to use and report (Q)SARs” Version 3.1 – July 2016



Weight of evidence OR supporting study

According to ECHA Practical guide “it should normally be a maximum of 2”
IUCLID includes several possibilities for explaining the assigned reliability.

Appropriate rationale should be chosen considering both VEGA AD and reliability evaluation and expert assessment

(Q)SAR results in IUCLID

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

Justification for type of information

1. SOFTWARE

VEGA v1.2.3

2. MODEL (incl. version number)

Mutagenicity ISS Model (version 1.0.3)

3. SMILES OR OTHER IDENTIFIERS USED AS INPUT FOR THE MODEL

c1ccc(cc1)c2ccc(c(c2Cl)Cl)Cl

4. SCIENTIFIC VALIDITY OF THE (Q)SAR MODEL

[[Explain how the model fulfils the OECD principles for (Q)SAR model validation. Consider attaching the QMRF and/or QPRF or providing a link]

- Defined endpoint:
- Unambiguous algorithm:
- Defined domain of applicability:
- Appropriate measures of goodness-of-fit and robustness and predictivity:
- Mechanistic interpretation:

(QMRF) can be attached (next slide) and referenced here

5. APPLICABILITY DOMAIN

[Explain how the substance falls within the applicability domain of the model]

- Descriptor domain:
- Structural domain:
- Mechanistic domain:
- Similarity with analogues in the training set:
- Other considerations (as appropriate):

VEGA report can be attached and used as reference. However, if expert assessment is performed, it can be described here.

procedure used to identify similar molecules in Danish QSAR database con be explained here

6. ADEQUACY OF THE RESULT

[Explain how the prediction fits the purpose of classification and labelling and/or risk assessment]

Expert assessment is needed

(Q)SAR results in IUCLID

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

Attached justification + New item | Import file

#	Attached justification	Reason / purpose	Actions
1	QMRF_MUTA_ISS.pdf	(Q)SAR model reporting (QMRF)	

QPRF can also be attached, if prepared by the user

Data source

Reference
VEGA v1.2.3 | 2023
Data access
data published

Basic information about the software and model are sufficient

Materials and methods

Test guideline + New item | Import file

#	Qualifier	Guideline	Version / remarks	Deviations	Actions
1	according to guideline	other: REACH Guidance on QSARs R.6	*None*	*None*	

Otherwise, the test guidelines used to generate the data for the training set

Principles of method if other than guideline

The Benigni / Bossa rulebase for mutagenicity and carcinogenicity - a module of Toxtree, (2008) by R. Benigni, C. Bossa, N. Jeliazkova, T. Netzeva, and A. Worth. European Commission report EUR 23241 EN

R. Benigni, C. Bossa, T. Netzeva, A. Rodomonte, and I. Tsakovska (2007) Mechanistic QSAR of aromatic amines: new models for discriminating between mutagens and nonmutagens, and validation of models for carcinogens. Environ mol mutag 48:754-771

Benfenati E, Manganaro A, Gini G. VEGA-QSAR: AI inside a platform for predictive toxicology Proceedings of the workshop "Popularize Artificial Intelligence 2013", December 5th 2013, Turin, Italy Published on CEUR Workshop Proceedings Vol-1107

Information from QMRF section 2.7 - Reference(s) to main scientific papers and/or software package

(Q)SAR results in IUCLID

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

Test material

Test material information
2,3,4-Trichlorobiphenyl | 2,3,4-trichlorobiphenyl | 1,2,3-trichloro-4-phenylbenzene | 55702-46-0

Additional test material information
None

Specific details on test material used for the study
SMILES: c1ccc(cc1)c2ccc(c(c2Cl)Cl)Cl

Specific details on test material used for the study (confidential)
None

Test material must reflect the evaluated structure

If multiple constituents are assessed for one substance, the Practical Guide suggest to prepare separate entries

Results and discussion

Test results + New item | Import file

#	Key result	Species / strain	Metabolic act...	Genotoxicity	Cytotoxicity / ...	Vehicle contr...	Untreated ne...	True negative...	Positive contr...	Actions
1	☐	*None*	*None*	*None*	*None*	*None*	*None*	*None*	*None*	

Additional information on results
None

Remarks on result
no mutagenic potential (based on QSAR/QSPR prediction)

Any other information on results incl. tables
None

In the results section, the remarks picklist includes (Q)SAR-specific items

Overall remarks, attachments

Overall remarks
None

CREDITS

Acknowledgement:

- Nadine Sickinger, Milena Gillwald and the knoell Academy team
- Giuseppa “Nelly” Raitano (Mario Negri Institute), Eva Bay Wedebye (Danish Technical University) and all partners of the LIFE CONCERT REACH project,
- Katarzyna Bucior and the QSAR team at knoell
- The speakers of today and of 31 May

CONCERTREACH
CONCERTING EXPERIMENTAL DATA AND IN SILICO MODELS FOR REACH

LIFE17 GIE/IT/000461

ISTITUTO DI RICERCHE FARMACOLOGICHE MARIO NEGRI · IRCCS

big chem

kode from data to knowledge

knoell

DTU

Sviluppo chimica

LIFE CONCERT REACH – Web-seminars on practical examples on using (Q)SAR for REACH

knoell
worldwide registration

www.knoell.com

Think globally, act locally

knoell worldwide registration

CONCERTREACH CONCERTING EXPERIMENTAL DATA AND IN SILICO MODELS FOR REACH

LIFE17 GIE/IT/000461

Appendix A

Importance of using multiple models for addressing the same endpoint

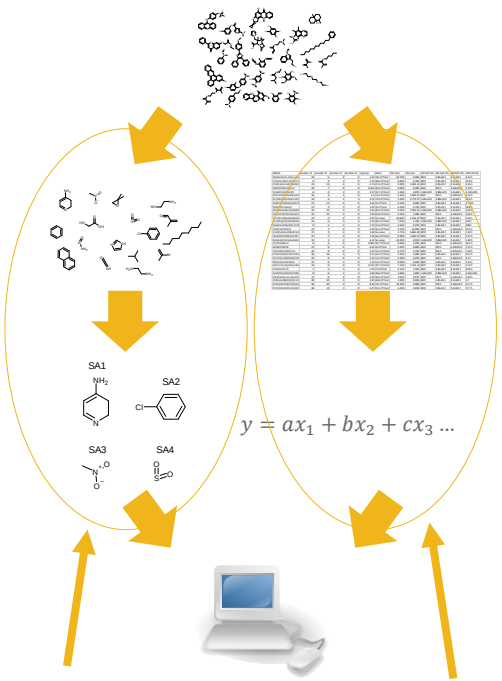
Development of statistical and knowledge-based models

knoell worldwide registration

CONCERTREACH CONCERTING EXPERIMENTAL DATA AND IN SILICO MODELS FOR REACH

LIFE17 GIE/IT/000461

Statistical models



Training set

Fragmentation and/or molecular descriptors

Identification of most relevant parameters

Implementation

VEGA
SarPy/IRFMN

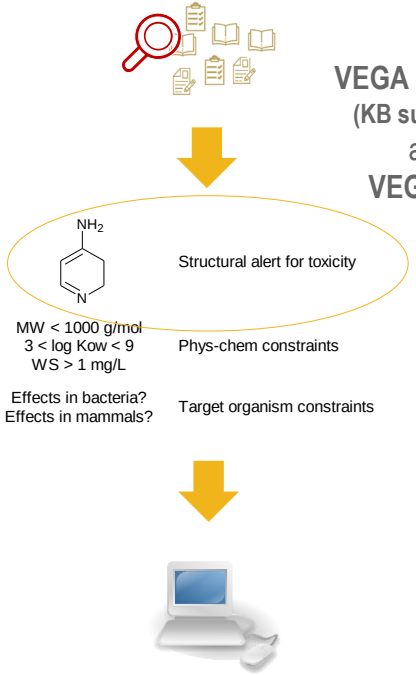
VEGA CAESAR
(Statistical submodel)

Knowledge-based models

Gathering expert knowledge

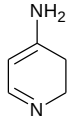
Rules development

Implementation



VEGA CAESAR
(KB submodel)
and
VEGA ISS

Statistical and KB models “complement” each other



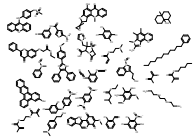
MW < 1000 g/mol
3 < log Kow < 9
WS > 1 mg/L

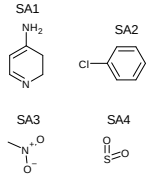
Effects in bacteria?
Effects in mammals?

Knowledge-based models

- Based on and **"limited"** by human knowledge;
- Support results with **literature and mechanistic information.**

First layer of integration





$$y = ax_1 + bx_2 + cx_3 \ldots$$

Statistical models

- Might be able to identify toxic effects **through mechanisms not known by expert**;
- **Can‘t support** such mechanism nor the final evaluation **with literature evidence or explanations**.

Statistical and KB models “complement” each other

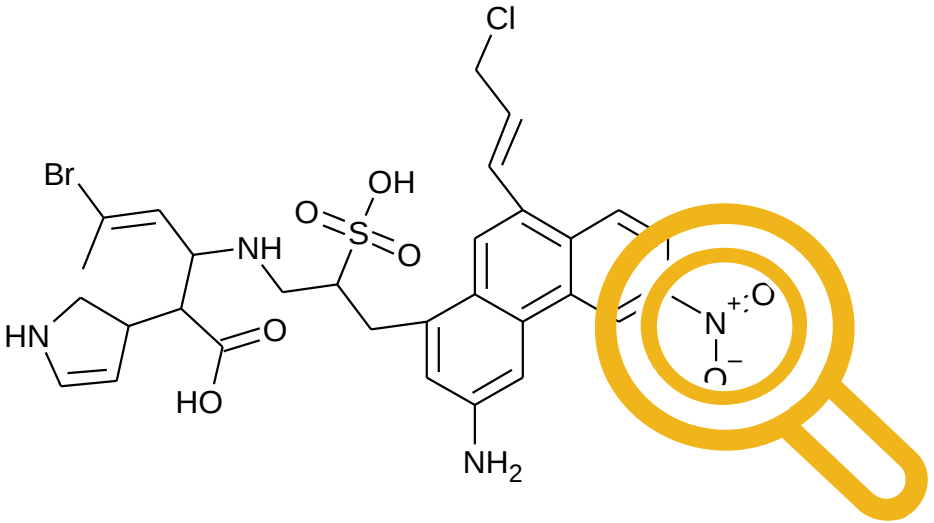
knoell worldwide registration

CONCERTREACH CONCERTING EXPERIMENTAL DATA AND IN SILICO MODELS FOR REACH

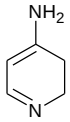
LIFE17 GIE/IT/000461

Second layer of integration

How the model “looks” at the structure to be predicted



Statistical and KB models: second integration layer



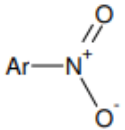
Knowledge-based models

- Focus on **specific substructures** (alerts);
- Might **not be able to evaluate potential influence of the structure on the toxicity** of the alert;
- Some models encode "exclusion rules", **still based on expert knowledge**:
  - Structural modifications that decrease or remove the toxicity;
  - Constraints on physico-chemical parameters.

E.g. Benigni-Bossa rulebase (ISS and CAESAR)

Ar = Any aromatic/heteroaromatic ring

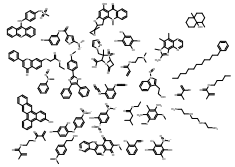
SA_27: Nitro-aromatic

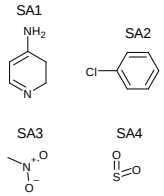


- Chemicals with ortho-disubstitution, or with an ortho carboxylic acid substituent are excluded.
- Chemicals with a sulfonic acid group (-SO3H) on the same ring of the nitro group are excluded .

Statistical and KB models: second integration layer

$$y = ax_1 + bx_2 + cx_3 \ldots$$



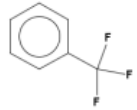


Statistical models

- Might be able to **consider the whole structure** for the toxicity prediction:
  - User evaluation **limited by "Black box"** effect;
- Might be able to **identify structural features that decrease the toxicity** of structural alerts:
  - **Critical expert evaluation** is needed.

E.g. SarPy/IRFMN model:
structural alert for non-mutagenicity

Fragment found: SM147



Sarpy alert n. 147 for NON-Mutagenicity, defined by the SMARTS: c1c(C(F)(F)F)cccc1

Following, the most similar compounds from the model's dataset having the same fragment.

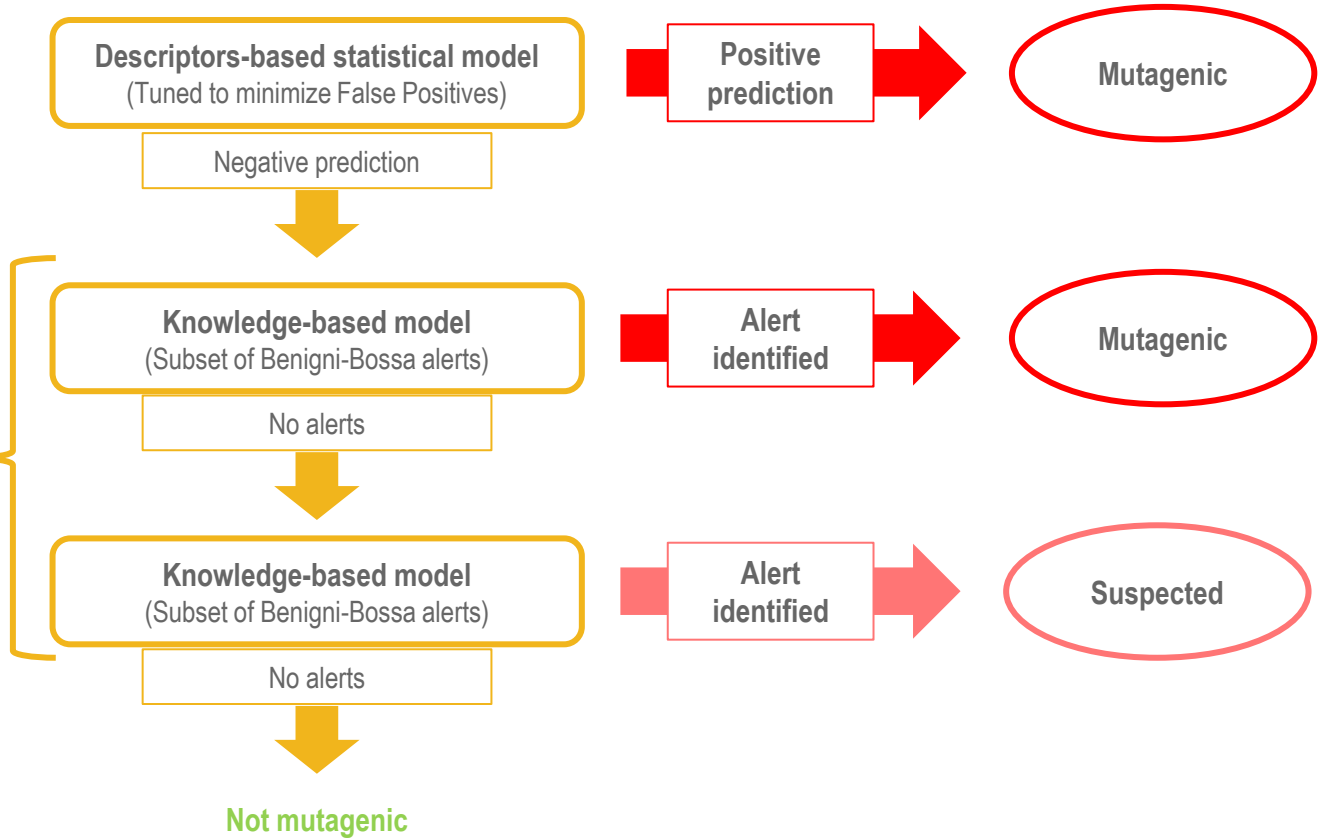
CAESAR model: Example of automated knowledge-based and statistical integration

knoell worldwide registration

CONCERTREACH CONCERTING EXPERIMENTAL DATA AND IN SILICO MODELS FOR REACH

LIFE17 GIE/IT/000461

Main aim: **identifying and "correcting" False Negative predictions**



knoell worldwide registration

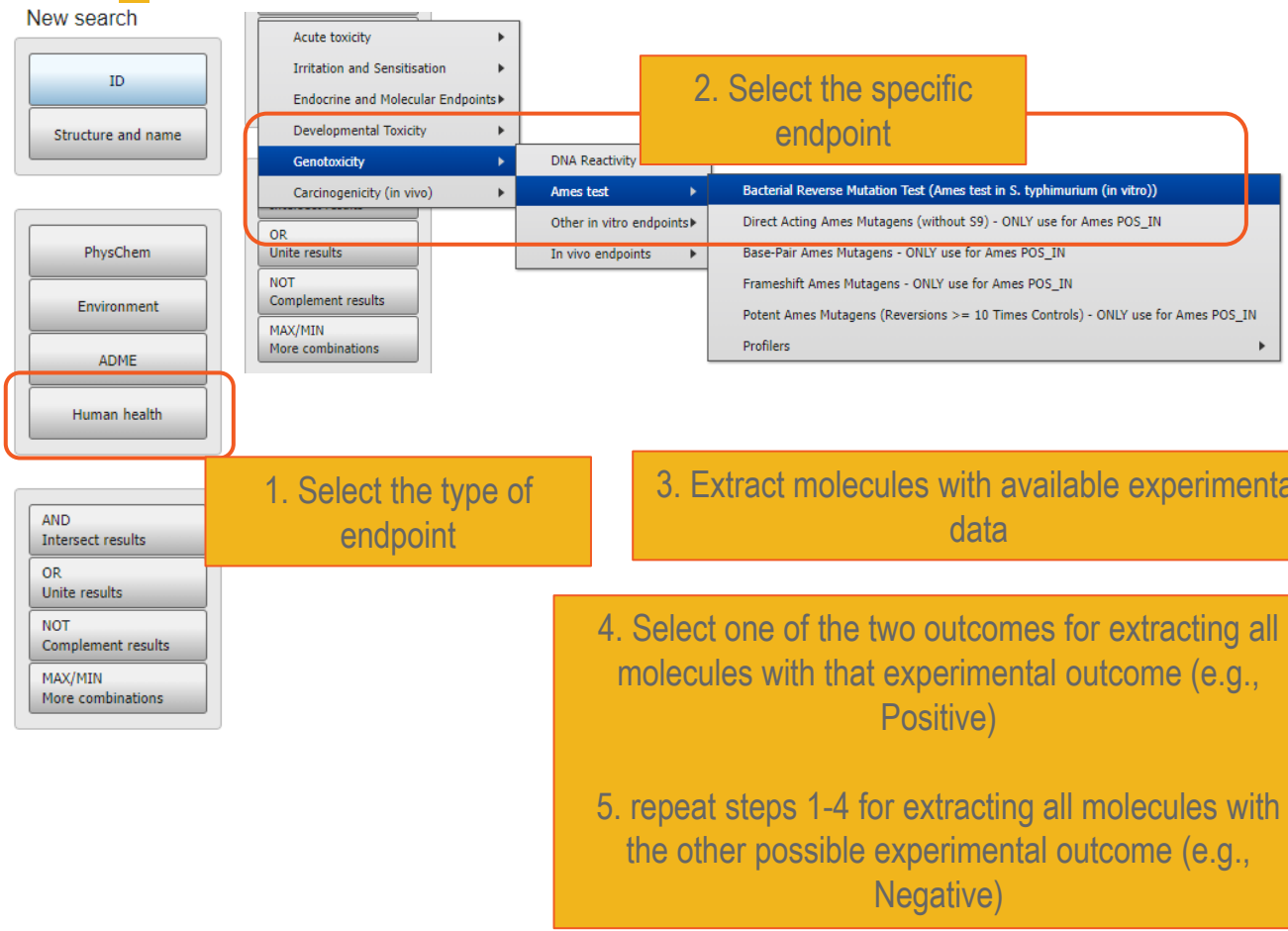
CONCERTREACH CONCERTING EXPERIMENTAL DATA AND IN SILICO MODELS FOR REACH

LIFE17 GIE/IT/000461

Appendix B

Danish QSAR Database: similar molecules identification

Danish (Q)SAR Database: identification of similar molecules



Danish (Q)SAR Database: identification of similar molecules

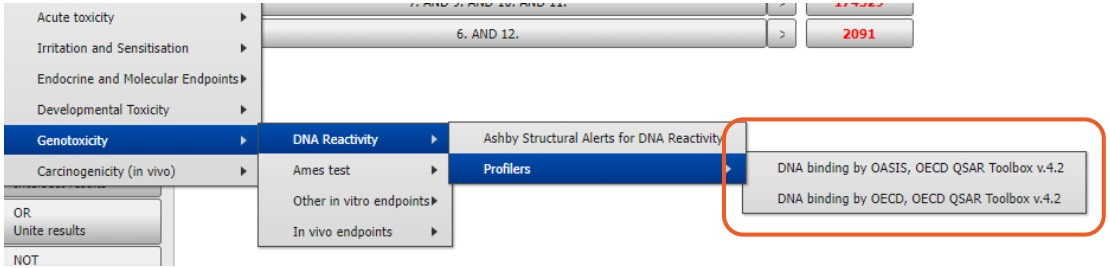
knoell worldwide registration

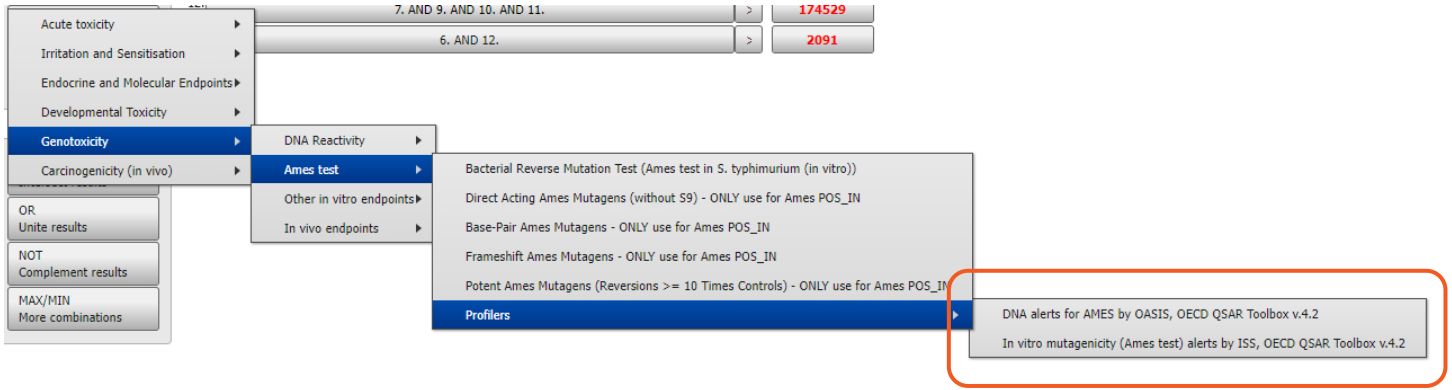
CONCERTREACH CONCERTING EXPERIMENTAL DATA AND IN SILICO MODELS FOR REACH

LIFE17 GIE/IT/000461

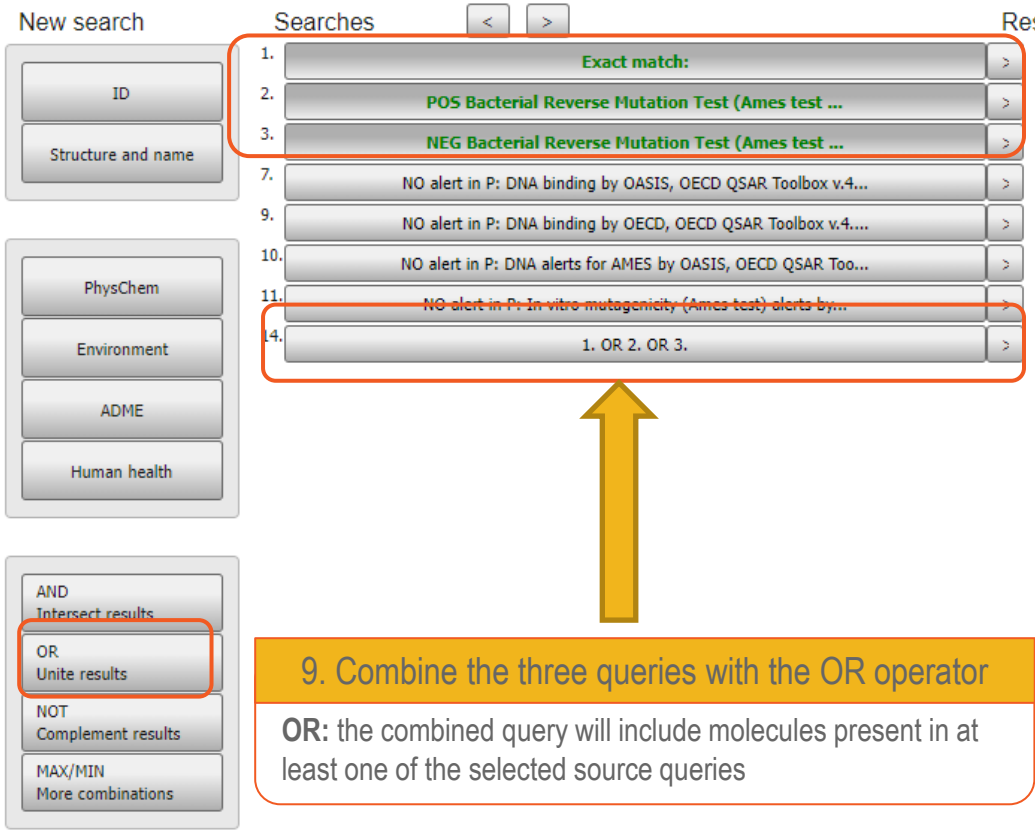
7. To account for mechanistic similarity, molecules with the same alerts profile can be extracted

In this case, according to the report generated for the target, no alerts for any of the profilers





Danish (Q)SAR Database: identification of similar molecules



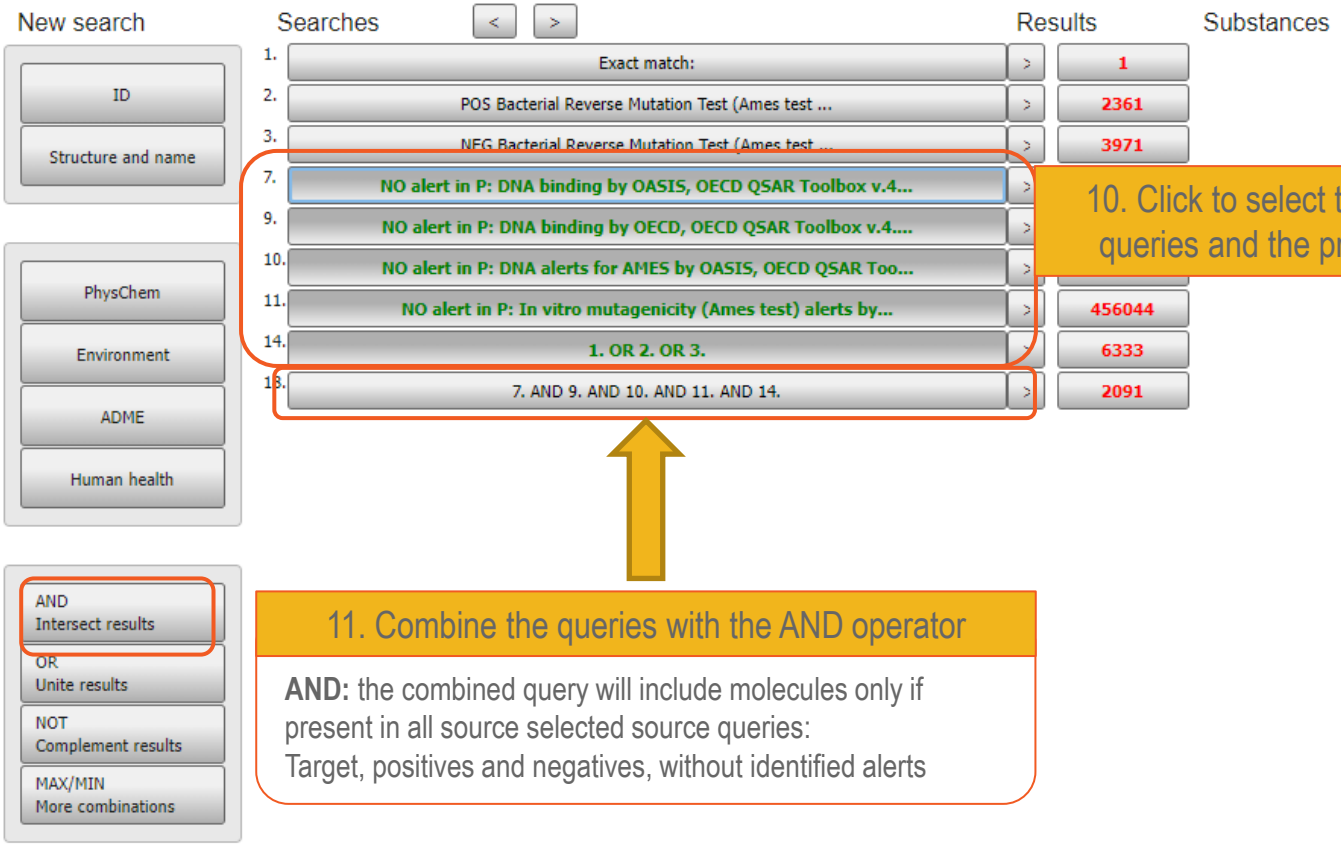
8. Click to select the results of all three queries

- **Target molecule**
- All molecule **experimentally positive** for Ames test
- All molecules **experimentally negative** for Ames test

9. Combine the three queries with the OR operator

OR: the combined query will include molecules present in at least one of the selected source queries

Danish (Q)SAR Database: identification of similar molecules



10. Click to select the results of “no alerts” queries and the previously combined list

11. Combine the queries with the AND operator

AND: the combined query will include molecules only if present in all source selected source queries:
Target, positives and negatives, without identified alerts