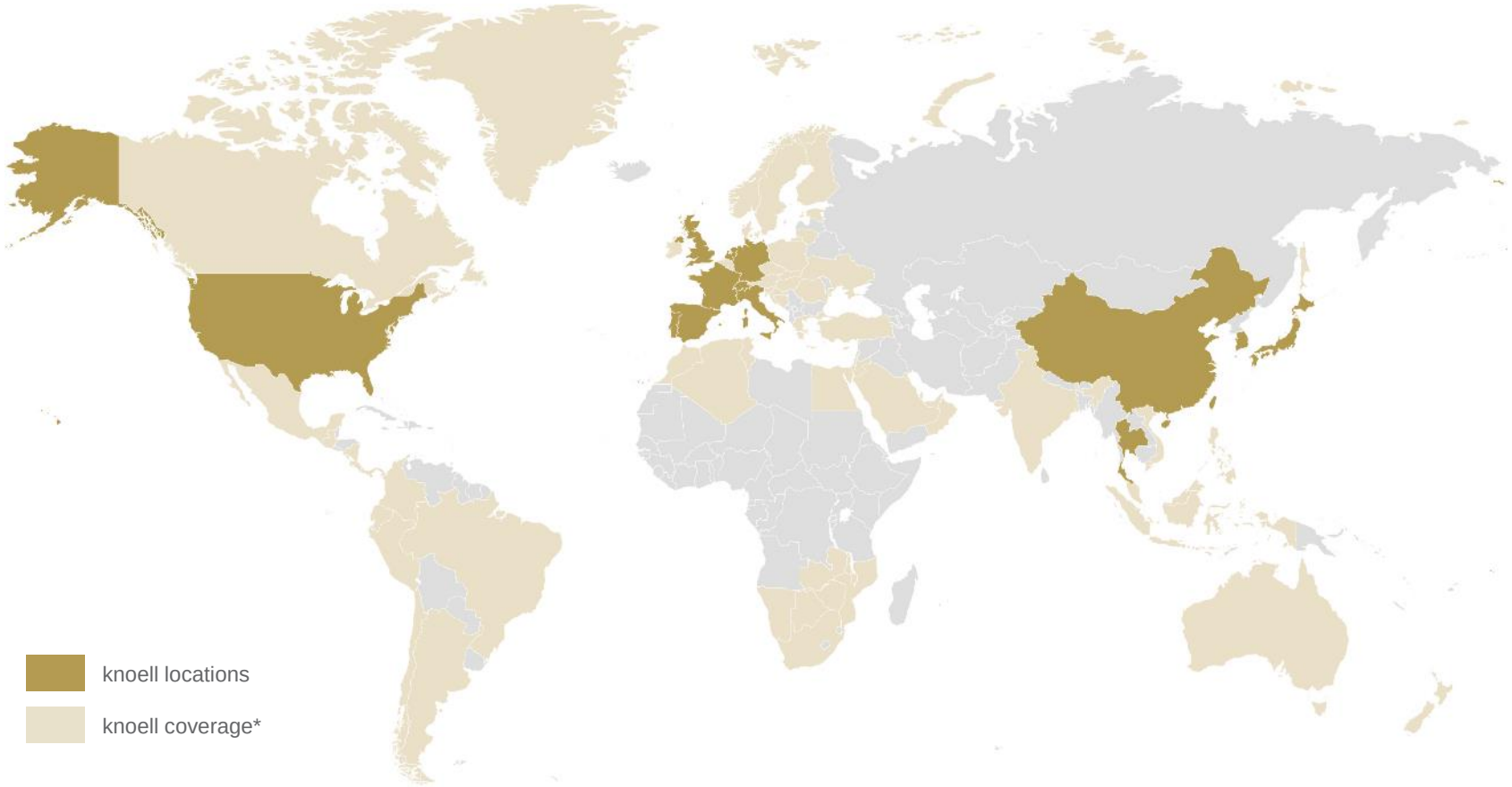
knoell
worldwide registration

Using QSAR results for REACH dossier preparation: lessons learned from a consultant company's perspective

19/06/2023, LIFE CONCERT REACH project – Final Workshop

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

knoell worldwide

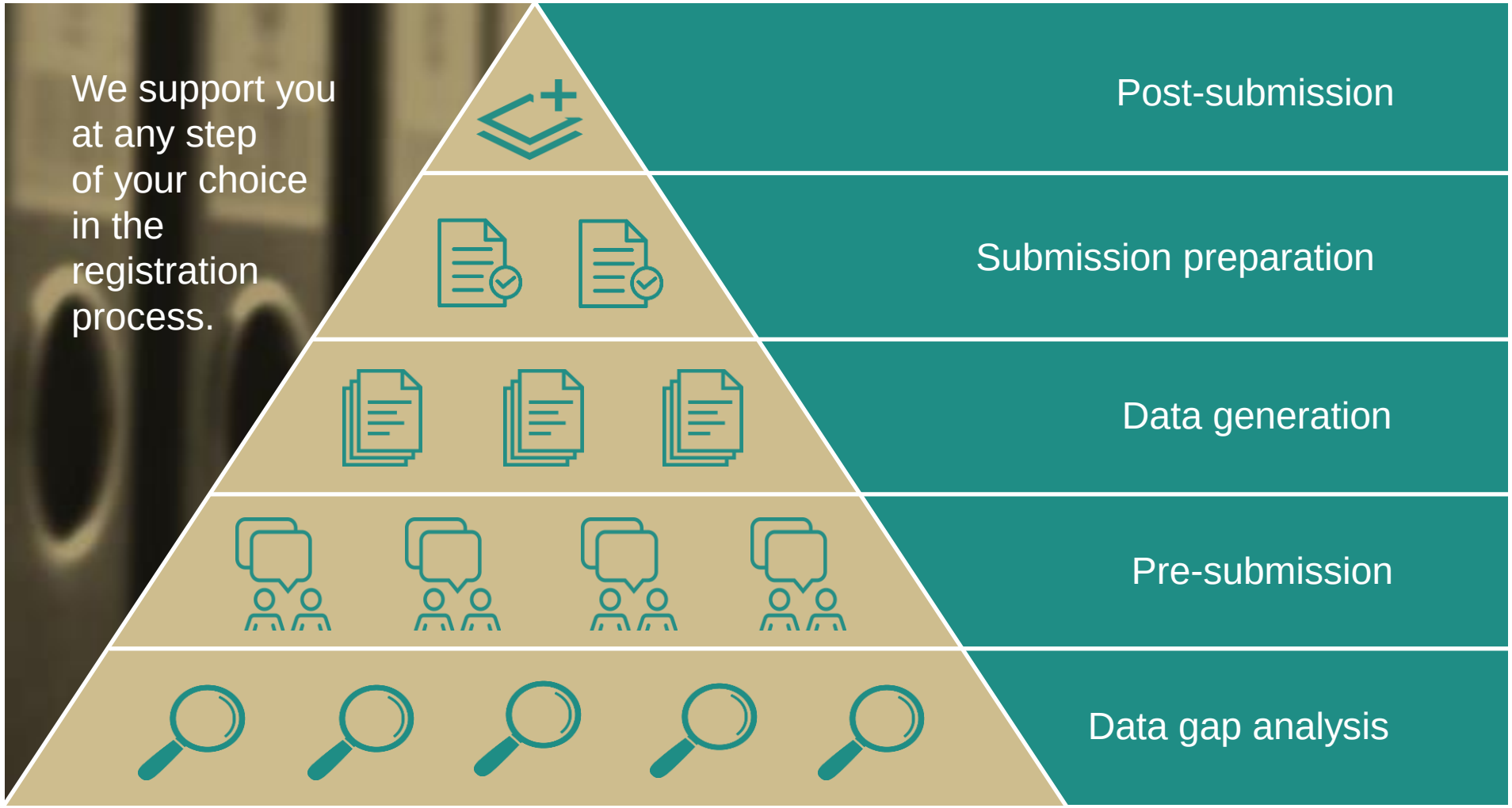


Our experts are located in North America, Europe and Asia with partners worldwide.

If a country is not covered by your interest, please contact us. We can make many things possible and are continuously expanding.

* service coverage can vary per country due to resources and expertise

Full-service or individual solutions



The knoell QSAR team

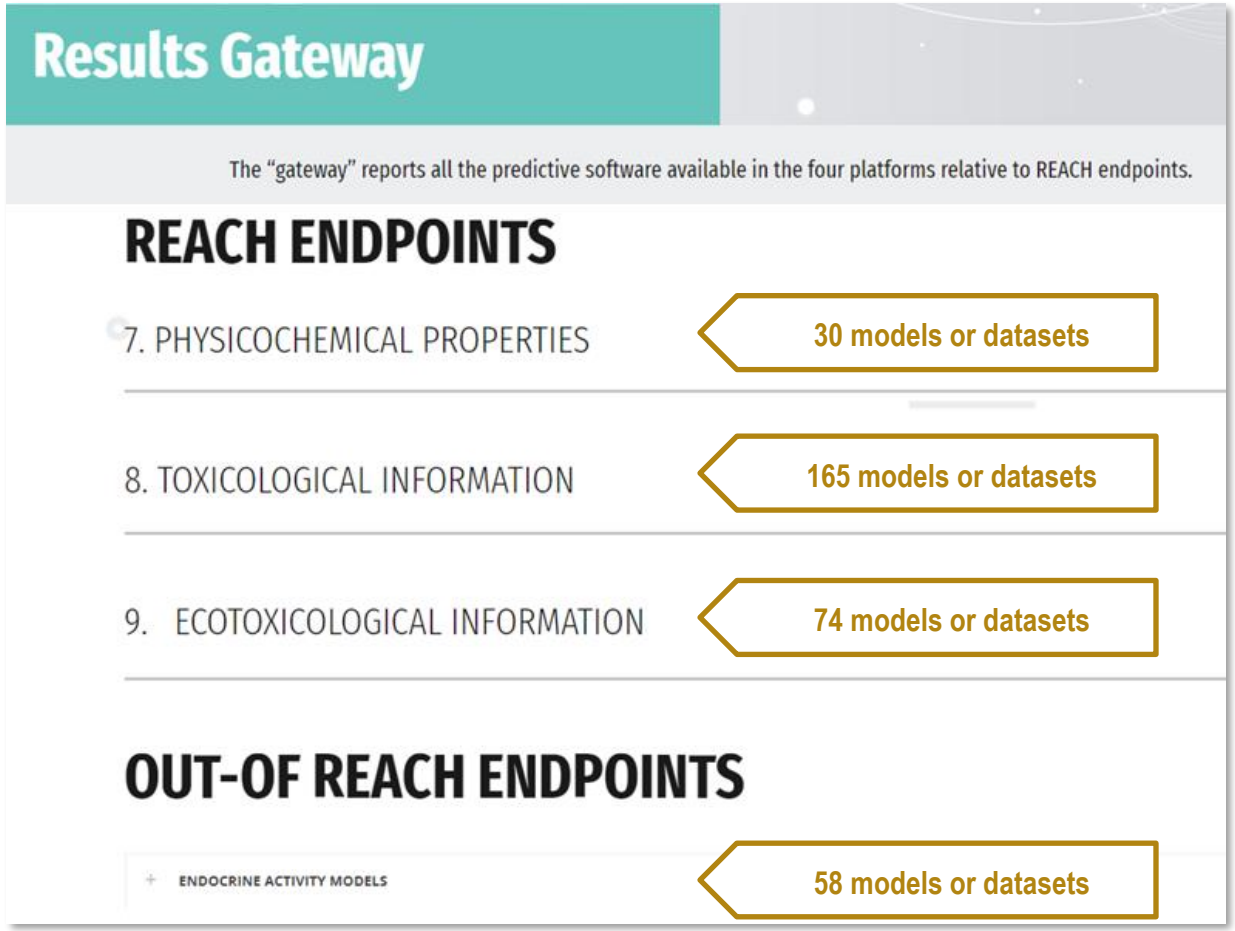
knoell worldwide registration

CONCERTREACH CONCERTING EXPERIMENTAL DATA AND IN SILICO MODELS FOR REACH

LIFE17 GIE/IT/000461

TABLE OF CONTENTS

In silico methods for regulatory purposes



More than 30 REACH endpoints addressed by:

- **234** statistical and knowledge-based **(Q)SAR models**;
- **35 sets of experimental data**.

The (Q)SAR models can be used within **other regulatory frameworks**.

In addition, **58 models** for evaluation of potential **endocrine activity**.

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

In silico methods for regulatory purposes

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
BCF	BCF (L/kg wet-wt) (EPI)	continuous					DanishQSARDatabase	
Log BCF	Log BCF (L/kg wet-wt) (EPI)	continuous					DanishQSARDatabase	
Whole Body Primary Biotransformation Fish Half-Life	Whole Body Primary Biotransformation Fish Half-Life (days) (EPI)	continuous					DanishQSARDatabase	
BCF Arnot-Gobas (upper trophic) Including Biotransformation	BCF Arnot-Gobas (upper trophic) Including Biotransformation (L/kg wet-wt) (EPI)	continuous					DanishQSARDatabase	
BCF Arnot-Gobas (upper trophic) Zero Biotransformation	BCF Arnot-Gobas (upper trophic) Zero Biotransformation (L/kg wet-wt) (EPI)	continuous					DanishQSARDatabase	
BAF Arnot-Gobas (upper trophic) Including Biotransformation	BAF Arnot-Gobas (upper trophic) Including Biotransformation (L/kg wet-wt) (EPI)	continuous					DanishQSARDatabase	
BAF Arnot-Gobas (upper trophic) Zero Biotransformation	BAF Arnot-Gobas (upper trophic) Zero Biotransformation (L/kg wet-wt) (EPI)	continuous					DanishQSARDatabase	
Bioaccumulation in								

For each endpoint, **multiple models are available**

Examples:
- 12 models for bioconcentration factor (BCF)
- 8 models for octanol/water partition coefficient (log Kow)
- 24 bacterial mutagenicity (Ames test) models

Endpoints selected as case studies for practical examples
Next section of this presentation

In silico methods for regulatory purposes

With all these models available, questions can arise:

Q1: Which model(s) should I use?

Q2: Do **regulators** indicate reliable models?

Q3: Which data can be generated and for **which purpose**?

In silico methods for regulatory purposes

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

Q1: Which model(s) should I use?
Q2: Do **regulators** indicate reliable models?

- ***A priori* selection is generally not possible**
  - Indeed, **regulators don't/can't give clear indications**
  - However, **experience in using the models and information from** developers (e.g., which substances compose the training set) **might suggest** which model could give more reliable results for certain type of substances (e.g., industrial chemicals, active substances, etc.)
- Selection can be based on:
  - 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

Tools in LIFE CONCERT REACH gateway provide the required information

(next section)

In silico methods for regulatory purposes

Q3: Which data can be generated and for **which purpose**?

REACH Regulation (EC) No 1907/2006

Results from (Q)SAR model predictions can be used for:

- Experimental data replacement (e.g., physico-chemical and environmental fate properties)
- Supporting (eco)toxicological data (e.g., weight of evidence (WoE))
- Impurities evaluation
- Supporting read-across strategies
- Screening and prioritization
- Testing strategies definition

In silico methods for regulatory purposes

Q3: Which data can be generated and for **which purpose**?

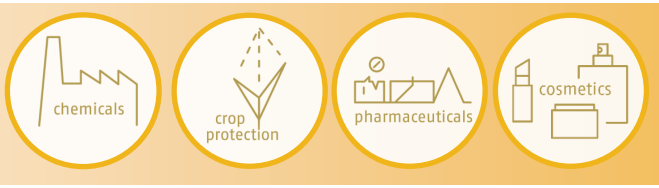
In silico methods are increasingly used in several other regulatory frameworks

Impurities (eco)toxicological assessment

(when **no experimental data available** or for **testing strategy**)

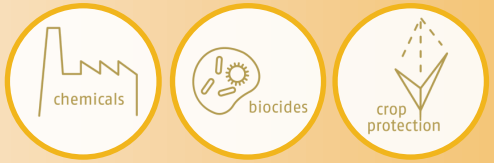
Grouping strategies

(for **read-across**)



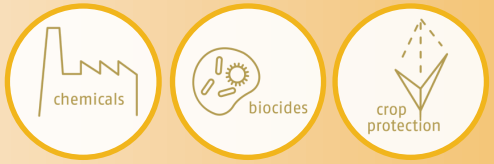
Metabolites or residues prediction or assessment

(when **no experimental data available** or for **testing strategy**)



Endocrine Disruption assessment

(**WoE** with experimental evidence)



Non-intentionally added substances (NIAS) assessment

(when **no experimental data available** or for **testing strategy**)

In silico methods for regulatory purposes

knoell worldwide registration

CONCERTREACH
CONCERTING EXPERIMENTAL DATA AND IN SILICO MODELS FOR REACH

LIFE17 GIE/IT/000461

Take home messages

- *In silico* models are **increasingly accepted** within several EU regulations
- CONCERT REACH gateway: **>300 models** for **30 REACH endpoints** and **potential endocrine activity** evaluation
- Regulators do not provide recommendation about the models to use, ***a priori* decision is not possible**
- For each endpoint, **multiple models should be used**
- Results from *in silico* models can **support**, **replace** or **fine tune experimental testing**

knoell worldwide registration

CONCERTREACH
CONCERTING EXPERIMENTAL DATA AND IN SILICO MODELS FOR REACH

LIFE17 GIE/IT/000461

TABLE OF CONTENTS

In silico models: practical examples of using models in the gateway

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

Four case studies have been presented during two web-seminars

17 May 2023 **VEGA CAESAR** model + **Danish QSAR database consensus** model for ***in vitro*** **gene mutation in bacteria**

VERA automated read-across and application on **carcinogenicity**

31 May 2023 **3 VEGA models** (Meylan/KOWWIN, ALogP and MLogP) for **octanol/water partition coefficient**

VEGA CAESAR model + **OCHEM Gramatica & Papa** model for **bioconcentration factor**

Aim: showing applications of **different models** from **different platforms** from the CONCERT REACH network

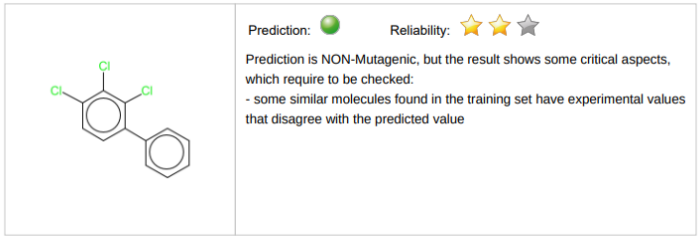
- **8 (Q)SAR** models from **3 platforms**, covering **3** REACH **endpoints**
- 1 novel **automated read-across** tool (VERA), applied on a **4th endpoint**

For (Q)SAR models, **preparation of IUCLID entries for REACH dossier** is demonstrated, according to ECHA's Practical guide - How to use and report (Q)SARs

Recordings and slides will be available on CONCERT REACH website

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

VEGA: Example of critical evaluation of the automated Applicability Domain (AD) / reliability evaluation



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

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

Compound #1
CAS: N.A.
Dataset id:441 (Training Set)
SMILES: c1cc(c(cc1c2ccc(c(c2Cl)Cl)Cl)Cl)Cl
Similarity: 0.925
Experimental value : NON-Mutagenic
Predicted value : NON-Mutagenic

Compound #2
CAS: 91-94-1
Dataset id:458 (Training Set)
SMILES: Nc1ccc(cc1Cl)c2ccc(N)c(c2)Cl
Similarity: 0.86
Experimental value : Mutagenic
Predicted value : Mutagenic

Alerts (not found also in the target): SA28 Primary aromatic amine, hydroxyl amine and its derived esters (with restrictions)

Compound #3
CAS: 72-54-8
Dataset id:473 (Training Set)
SMILES: c1cc(ccc1C(c2ccc(cc2)Cl)C(Cl)Cl)Cl
Similarity: 0.828
Experimental value : NON-Mutagenic
Predicted value : Mutagenic

Alerts (not found also in the target): SA8 Aliphatic halogens

Compound #4
CAS: 72-55-9
Dataset id:176 (Training Set)
SMILES: c1cc(ccc1C(c2ccc(cc2)Cl)=C(Cl)Cl)Cl
Similarity: 0.815
Experimental value : NON-Mutagenic
Predicted value : NON-Mutagenic

Compound #5
CAS: 50-29-3
Dataset id:751 (Training Set)
SMILES: c1cc(ccc1C(c2ccc(cc2)Cl)C(Cl)(Cl)Cl)Cl
Similarity: 0.813
Experimental value : NON-Mutagenic
Predicted value : NON-Mutagenic

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**

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.

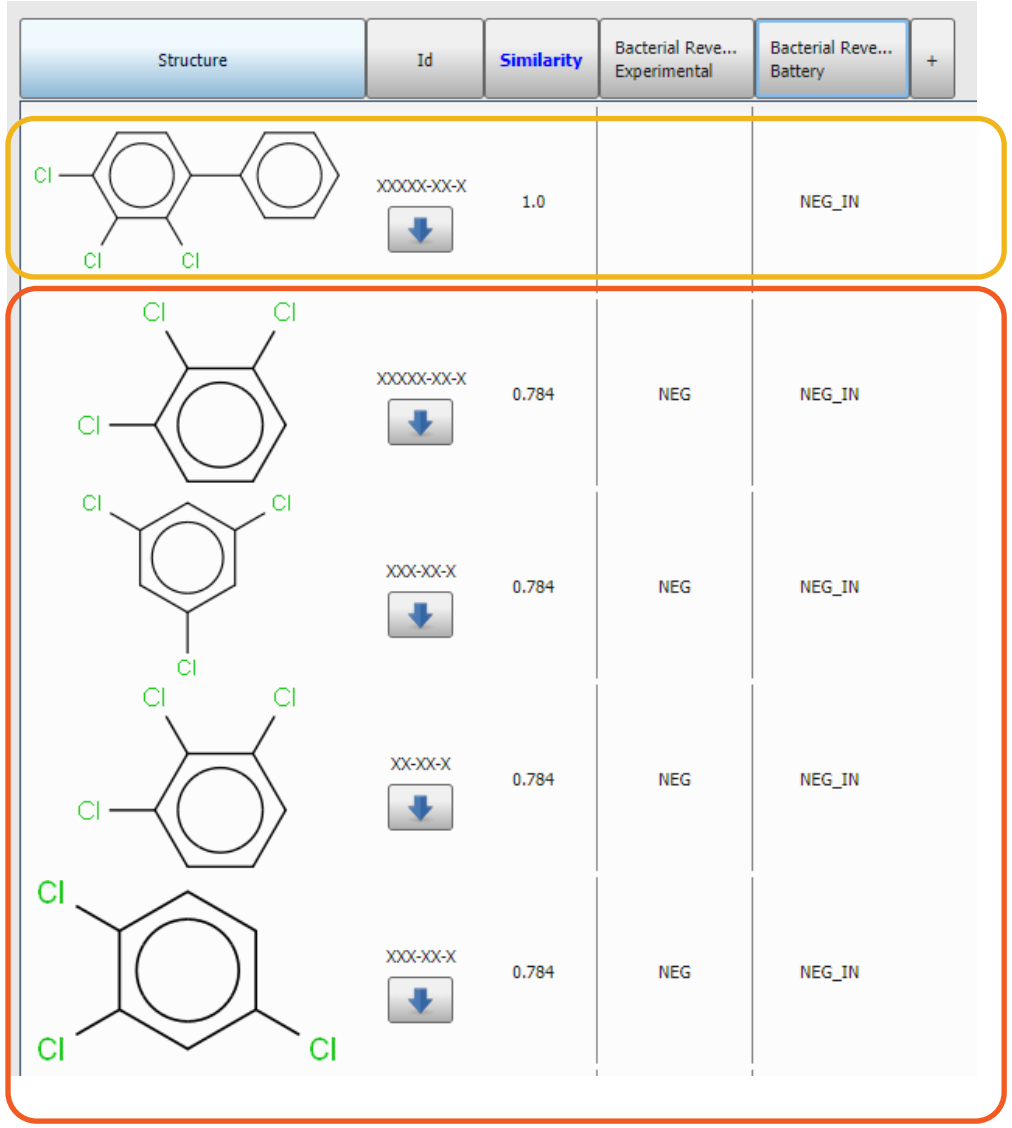
Structural alerts identified by two endpoint-specific profilers present in the OECD QSAR Toolbox

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



Structure	Id	Similarity	Bacterial Reve... Experimental	Bacterial Reve... Battery	+
	XXXXX-XX-X	1.0		NEG_IN	
	XXXXX-XX-X	0.784	NEG	NEG_IN	
	XXX-XX-X	0.784	NEG	NEG_IN	
	XX-XX-X	0.784	NEG	NEG_IN	
	XXX-XX-X	0.784	NEG	NEG_IN	

Target molecule

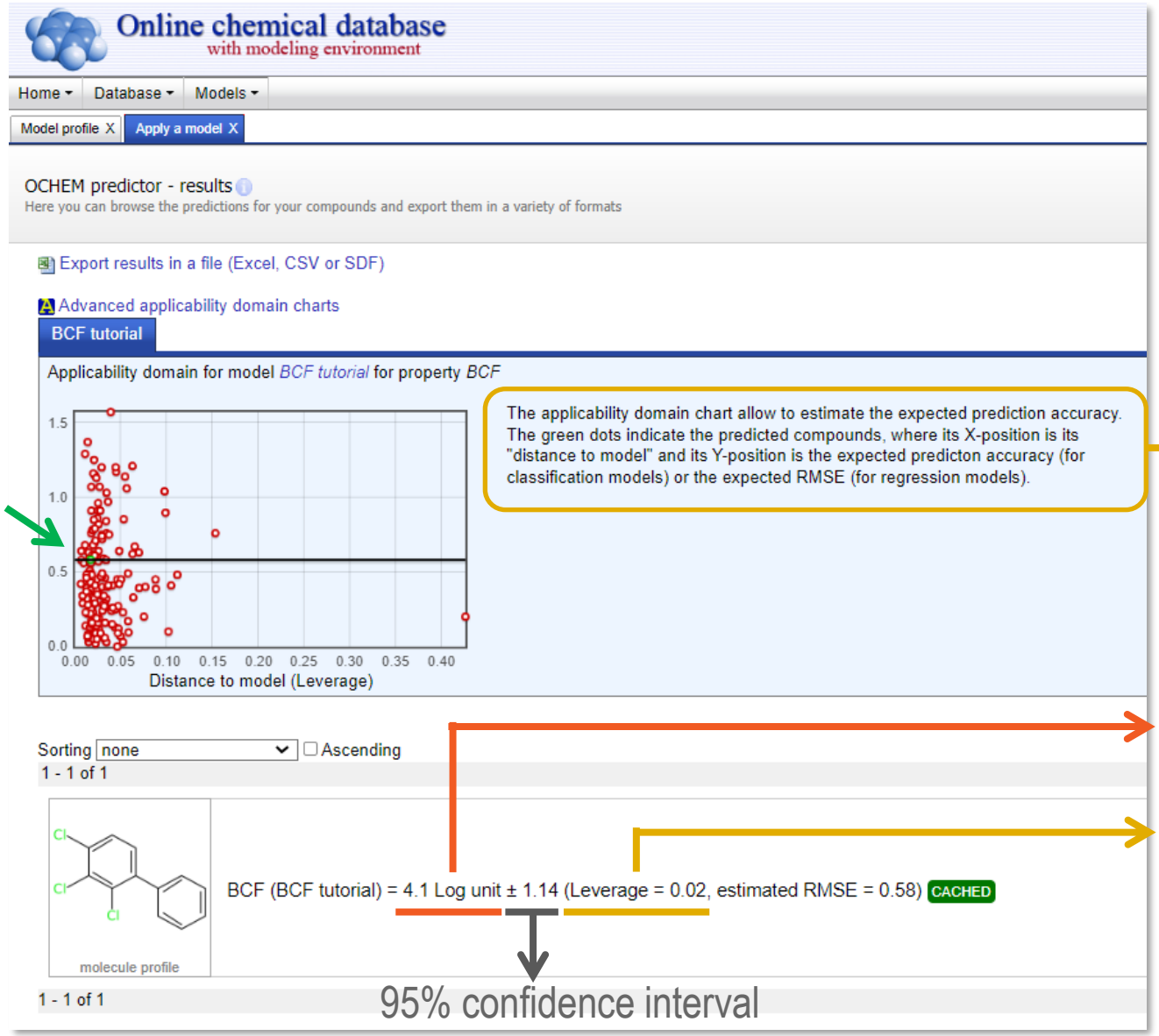
Supporting similar molecules

- **Not automatically** provided
- **Manual** 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**.

OCHEM: results and AD for bioconcentration factor



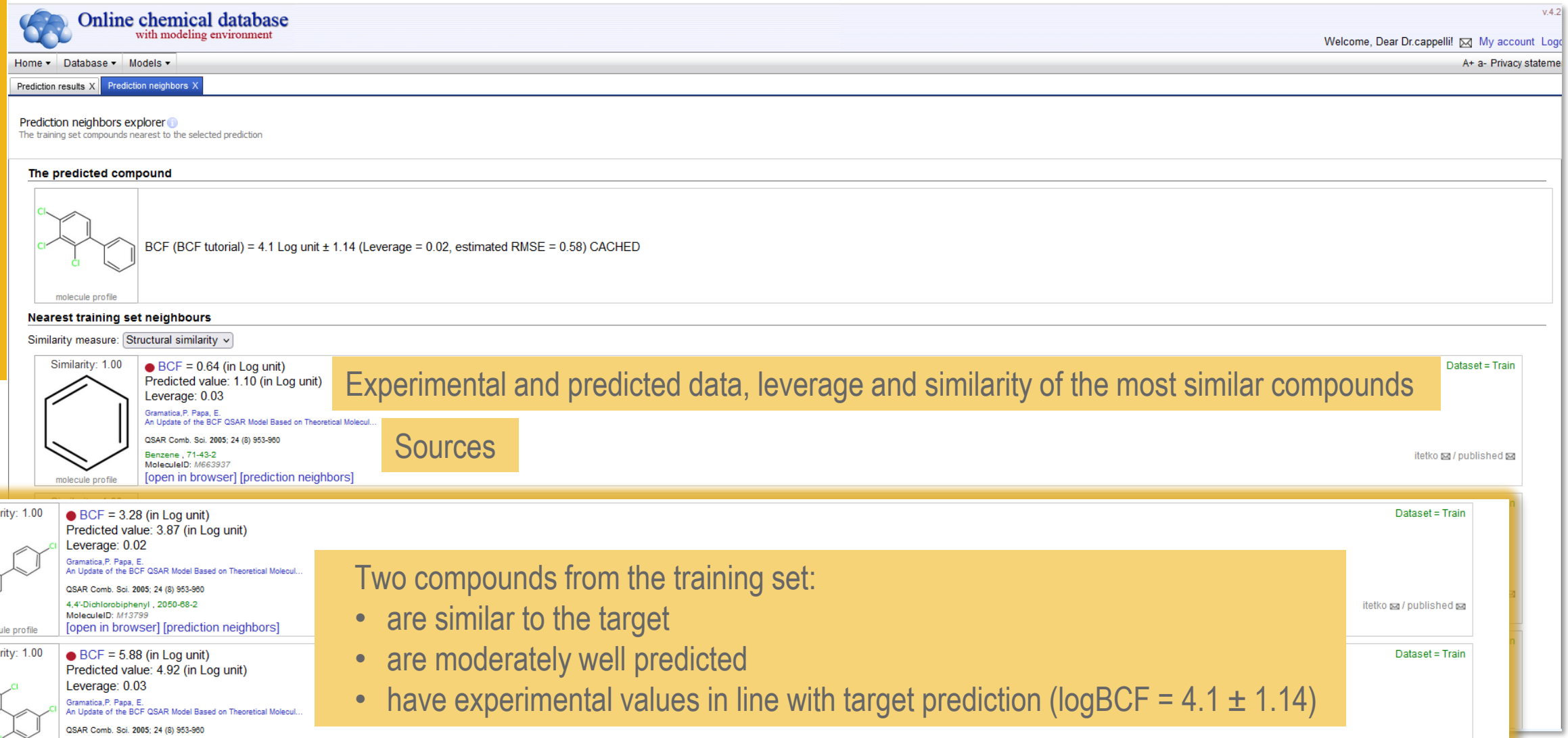
Target chemical compliant with the AD of the model

Compound predicted as bioaccumulative (LogBCF > 3.3)

"Distance to model" value

OCHEM: similar molecules as supporting information

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



In silico models: practical examples

knoell worldwide registration

CONCERTREACH
CONCERTING EXPERIMENTAL DATA AND IN SILICO MODELS FOR REACH

LIFE17 GIE/IT/000461

Take home messages

- The *in silico* tools integrated in the gateway **provide information for expert evaluation** of the generated results and associated reliability:
  - VEGA, OCHEM and Danish QSAR database: **automated applicability domain** evaluation;
  - VEGA and OCHEM: **automated** extraction of **similar molecules**;
  - Danish QSAR database: non-automated but **"customizable" similar molecules** identification.

- Fore more information: **recording and slides of the case studies will be published** on the LIFE CONCERT REACH project website.

knoell worldwide registration

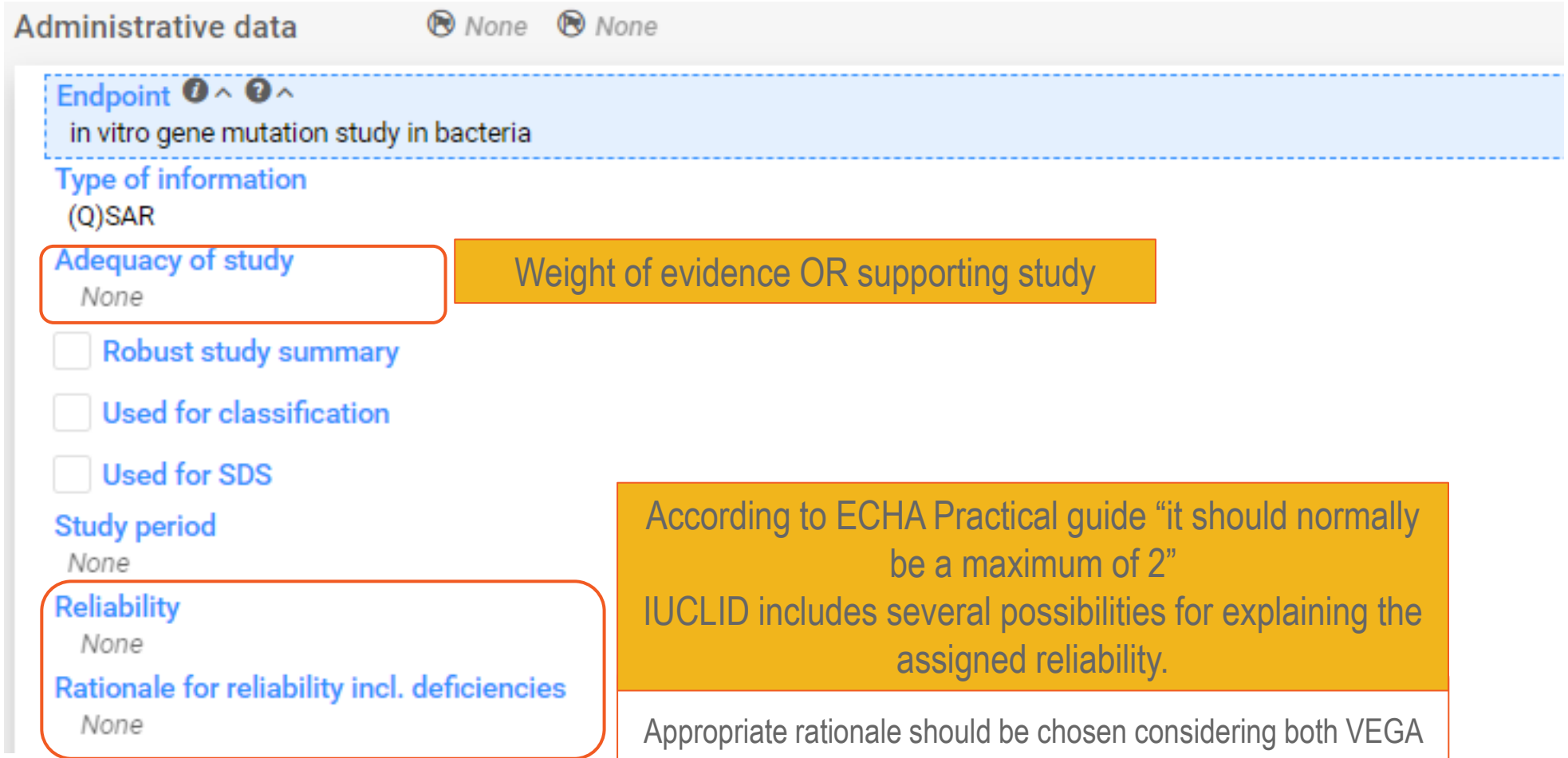
CONCERTREACH
CONCERTING EXPERIMENTAL DATA AND IN SILICO MODELS FOR REACH

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 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 can 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

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

CREDITS

Acknowledgement:

- Katarzyna Bucior, Antje Gerloff-Elias and the QSAR team at knoell
- All partners of the LIFE CONCERT REACH project

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

Using QSAR results for REACH dossier preparation: lessons learned from a consultant company's perspective

knoell
worldwide registration