SOT | Society of Toxicology

Creating a Safer and Healthier World by Advancing the Science and Increasing the Impact of Toxicology

Computational Toxicology Specialty Section

CTSS Webinar Series

Scheduled Webinars on Computational Toxicology Topics

2019-2020

CTSS Executive Committee

President
Lisa Beilke PhD DABT ERT
Toxicology Solutions

Vice-President
David Szabo PhD DABT
PPG

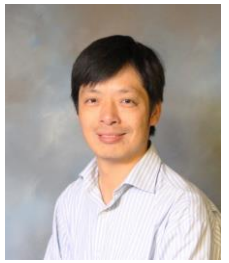
Vice-Pres Elect
Catrin Hasselgren PhD
Genentech

Past President
David Woolley PhD DABT ERT
ForthTox

Treasurer
Mark Gosink PhD
Pfizer

Secretary
Joel Cohen ScD DABT
Gradient

Councilor
Nigel Greene PhD
Astra Zeneca



Councilor
Minjun Chen PhD
FDA

Student Rep
Linlin Zhao PhD cand.
Rutgers University

SOT | Society of Toxicology

Creating a Safer and Healthier World by Advancing the Science and Increasing the Impact of Toxicology

Computational Toxicology Specialty Section

2019-2020 CTSS Webinar Series:

An introduction to *in silico* toxicology

Presented by: Donna S. Macmillan (Lhasa Limited) & Glenn J. Myatt (Leadscope, Inc.)

Lhasa Limited

Leadscope®

Overview

- Introduction to different methodologies of *in silico* toxicology
- Expert rule-based methodologies
- Statistical-based methodologies
- Expert review and predicting toxicity
- Key learnings and conclusions

Introduction to different methodologies of *in silico* toxicology

Computational Toxicology
Specialty Section

An introduction...

Toxicology

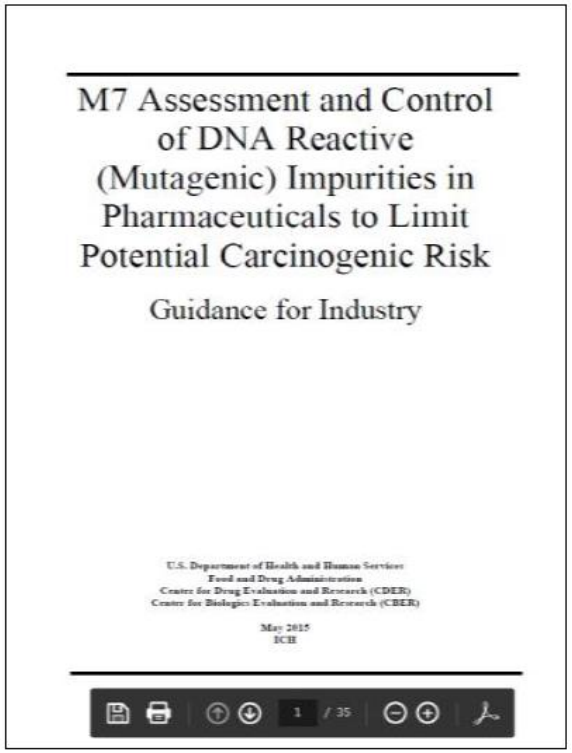
- ***Definition***: the study of the adverse health effects of chemicals on living organisms
- Often assessed using *in vivo* (animal) models
- However, *in vivo* tests are expensive, time-consuming, and are less desirable from an ethical standpoint
- Non-animal alternatives are increasingly being used in the pharmaceutical, pesticide, and cosmetic industries e.g. *in silico* toxicology

In silico toxicology

- ***Definition***: the development of mathematical and computer-based models to better understand and predict the adverse health effects of chemicals on living organisms
- No animals used
- *In silico* toxicology is fast, cheap, and can be used for multiple endpoints simultaneously
- Commonly used methodologies include **expert rule-based systems** and **statistical models**

Some common uses for *in silico* toxicology

- Assessment of impurities and degradation products
- Assessment of extractables and leachables
- Worker safety
- Selection of product development candidates



M7 Assessment and Control of DNA Reactive (Mutagenic) Impurities in Pharmaceuticals to Limit Potential Carcinogenic Risk

Guidance for Industry

U.S. Department of Health and Human Services
Food and Drug Administration
Center for Drug Evaluation and Research (CDER)
Center for Biologics Evaluation and Research (CBER)

May 2015
ICH

1 / 35

...an assessment of Structure-Activity Relationships (SAR) that focuses on **bacterial mutagenicity predictions** should be performed...

Two (Q)SAR prediction methodologies ... expert rule-based ... statistical-based.

(Q)SAR models utilizing these prediction methodologies should follow the general validation principles set forth by the Organisation for Economic Co-operation and Development (OECD).

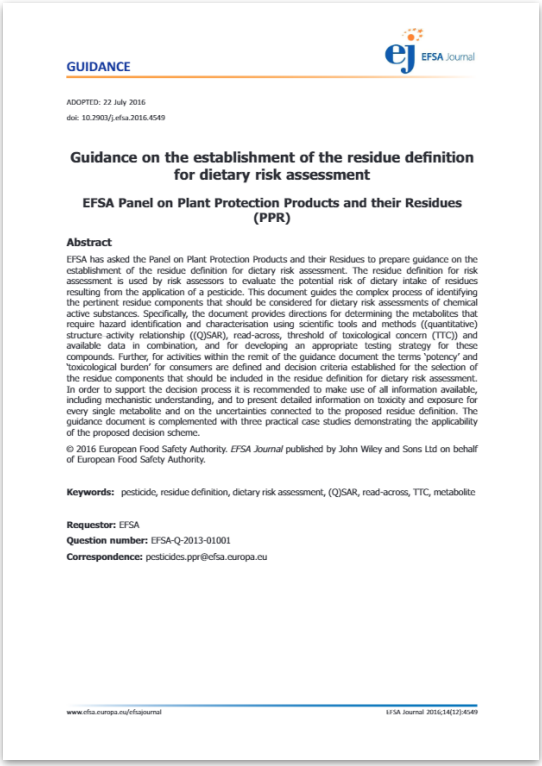
Expert rule-based model

Statistical-based model

2.3. Application of (Q)SAR and read-across for the prediction of genotoxicity (steps 5 and 6)

The genotoxicity assessment should be assisted by application of (Q)SAR (step 5 of the decision scheme, Figure 1) and read-across of metabolites (step 6) and could be optionally complemented in step 7 (see Section 2.4) with the consideration of exposure against the threshold of toxicological concern (TTC) for genotoxicity (0.0025 µg/kg bw per day). For compounds grouped according to their predicted effect, it is considered appropriate to apply the dose addition approach (EFSA Scientific Committee, 2012a; EFSA PPR Panel, 2013, 2014).

Step 5 includes the use of scientifically valid (Q)SAR models (see Section 2.3.1). The use of computational models for predictions of genotoxicity should not be based on the use of any single model alone. To maximise the sensitivity and specificity of the prediction, at least two independent (Q) SAR models, where possible, (e.g. based on different training sets and/or algorithms as knowledge-based and statistical-based models) should be applied for each genotoxicity endpoint (Worth et al., 2010, 2011a). A 'weight of evidence' approach should be followed, using all available information provided by the models, e.g. applicability domain, proposed mechanistic information, prediction for the similar substance.



GUIDANCE

EFSA Journal

ADOPTED: 22 July 2016

doi: 10.2903/j.efsa.2016.4549

Guidance on the establishment of the residue definition for dietary risk assessment

EFSA Panel on Plant Protection Products and their Residues (PPR)

Abstract

EFSA has asked the Panel on Plant Protection Products and their Residues to prepare guidance on the establishment of the residue definition for dietary risk assessment. The residue definition for risk assessment is used by risk assessors to evaluate the potential risk of dietary intake of residues resulting from the application of a pesticide. This document guides the complex process of identifying the pertinent residue components that should be considered for dietary risk assessments of chemical active substances. Specifically, the document provides directions for determining the metabolites that require hazard identification and characterisation using scientific tools and methods ((quantitative) structure activity relationship ((Q)SAR), read-across, threshold of toxicological concern (TTC)) and available data in combination, and for developing an appropriate testing strategy for these compounds. Further, for activities within the remit of the guidance document the terms 'potency' and 'toxicological burden' for consumers are defined and decision criteria established for the selection of the residue components that should be included in the residue definition for dietary risk assessment. In order to support the decision process it is recommended to make use of all information available, including mechanistic understanding, and to present detailed information on toxicity and exposure for every single metabolite and on the uncertainties connected to the proposed residue definition. The guidance document is complemented with three practical case studies demonstrating the applicability of the proposed decision scheme.

Keywords: pesticide, residue definition, dietary risk assessment, (Q)SAR, read-across, TTC, metabolite

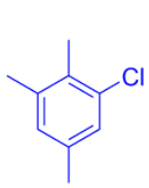
Requestor: EFSA

Question number: EFSA-Q-2013-01001

Correspondence: pesticides.ppr@efsa.europa.eu

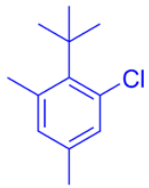
(Q)SAR models

- (Quantitative) Structure Activity Relationship
  - Identifies the general relationship between a chemical structure and the corresponding biological activity
  - Based on the assumption that *similar chemicals have similar physicochemical and biological properties*



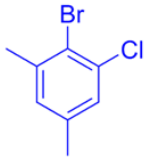
mp: 284.59 [K]
logP: 4.05
tPSA: 0

Toxicity: -



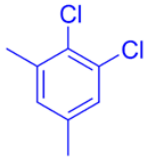
mp: 320.82 [K]
logP: 5.27
tPSA: 0

Toxicity: -



mp: 333.12 [K]
logP: 4.4
tPSA: 0

Toxicity: +



mp: 303.24 [K]
logP: 4.12
tPSA: 0

Toxicity: +

\+ = positive
\- = negative
mp = melting point
logP = how 'fat-soluble' the chemical is
tPSA = how permeable it is to cell membranes

SAR

Qualitative relationship between structure and biological activity (e.g. expert rule-based model)

QSAR

Mathematical (quantitative) relationship between structural features and/or physicochemical parameters and biological activity (e.g. statistical-based model)

(Q)SAR

Examples of (Q)SARs

Name	Developer	Expert / statistical?	Free / commercial?
CASE Ultra	MultiCASE Inc.	Statistical	Commercial
Derek Nexus	Lhasa Limited	Expert rule-based	Commercial
Leadscope Model Applier Expert Alerts	Leadscope Inc	Expert rule-based	Commercial
Leadscope Model Applier Statistical QSAR	Leadscope Inc	Statistical	Commercial
OECD QSAR Toolbox	OASIS, LMC	Expert + statistical	Free
Sarah Nexus	Lhasa Limited	Statistical	Commercial
TIMES-SS	OASIS, LMC	Expert + statistical	Commercial
TOPKAT	BIOVIA	Statistical	Commercial
ToxTree	IdeaConsult Ltd.	Expert	Free
Vega	Istituto di Ricerche Farmacologiche Mario Negri Milano	Statistical	Free

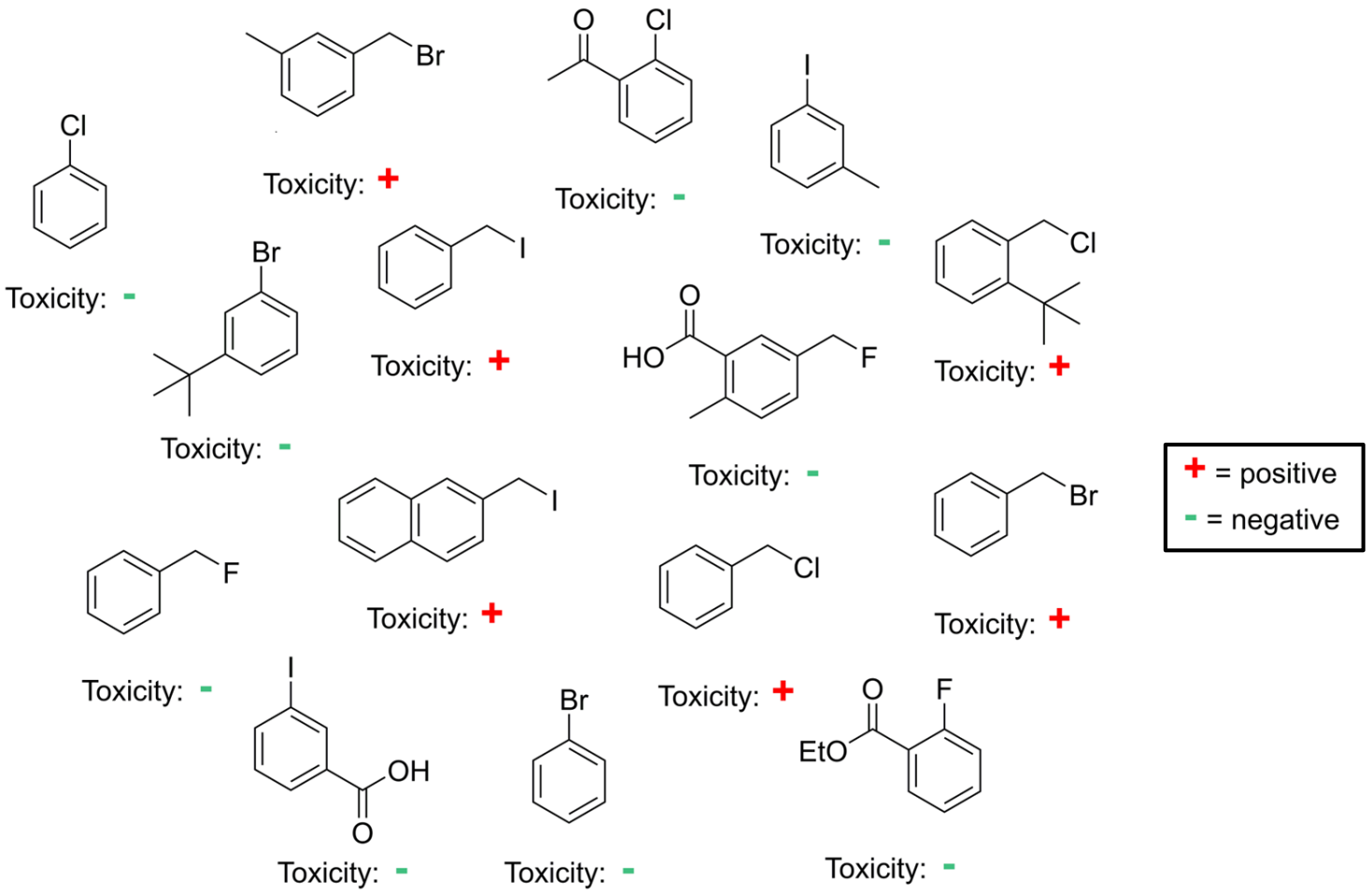
Expert rule-based methodologies

Computational Toxicology
Specialty Section

Expert rule-based methodologies

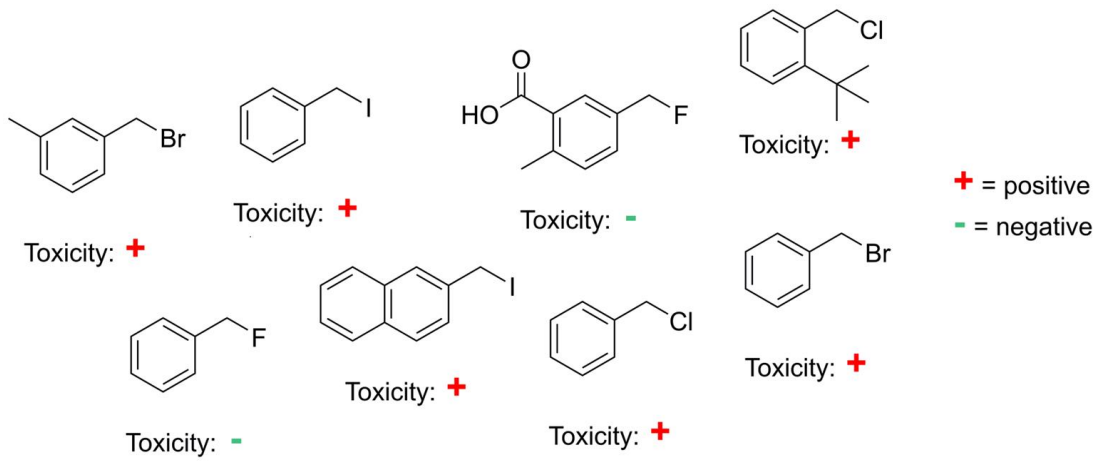
- Clusters of similar chemicals are extracted from public (or proprietary) datasets

Dataset

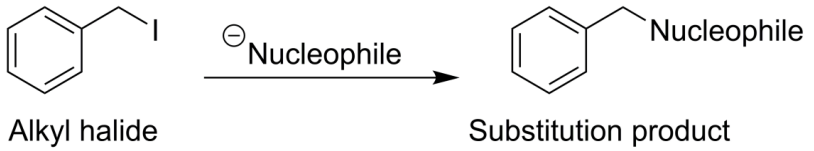


Expert rule-based methodologies

- Clusters of similar chemicals are extracted from public (or proprietary) datasets
- Experts use their knowledge to ascertain whether a structural alert can be identified for a specific cluster and provide a mechanistic justification



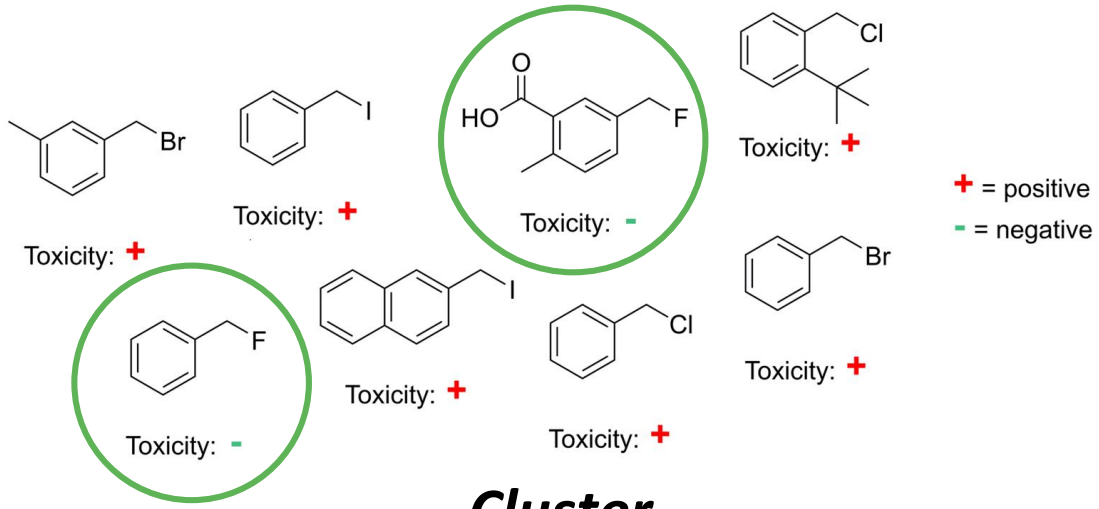
Cluster



Mechanism

Expert rule-based methodologies

- Clusters of similar chemicals are extracted from public (or proprietary) datasets
- Experts use their knowledge to ascertain whether a structural alert can be identified for a specific cluster and provide a mechanistic justification
- The SAR is rationalised and activating or deactivating features captured



Cluster

Compounds with fluorine substituents are associated with a lack of toxicity

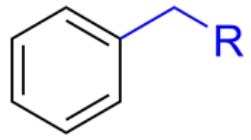
Rationalisation:

- Poorer leaving group
- Less reactive than other halogens

Structure-activity relationship (SAR)

Expert rule-based methodologies

- Clusters of similar chemicals are extracted from public (or proprietary) datasets
- Experts use their knowledge to ascertain whether a structural alert can be identified for a specific cluster and provide a mechanistic justification
- The SAR is rationalised and activating features or deactivating captured
- The SAR is encoded as a pattern to capture the toxicophore(s)

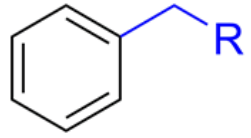


R = Cl, Br, I

Toxicophore (pattern) encoded in alert

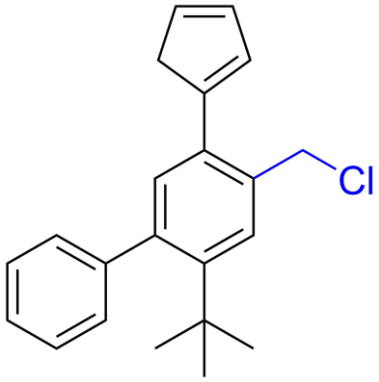
Expert rule-based methodologies

- Clusters of similar chemicals are extracted from public (or proprietary) datasets
- Experts use their knowledge to ascertain whether a structural alert can be identified for a specific cluster and provide a mechanistic justification
- The SAR is rationalised and activating features or deactivating captured
- The SAR is encoded as a pattern to capture the toxicophore(s)
- If a query compound includes the toxicophore, an alert is fired

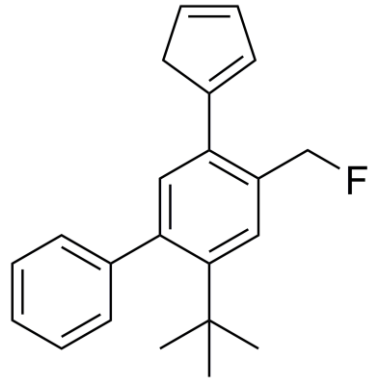


R = Cl, Br, I

Toxicophore (pattern) encoded in alert

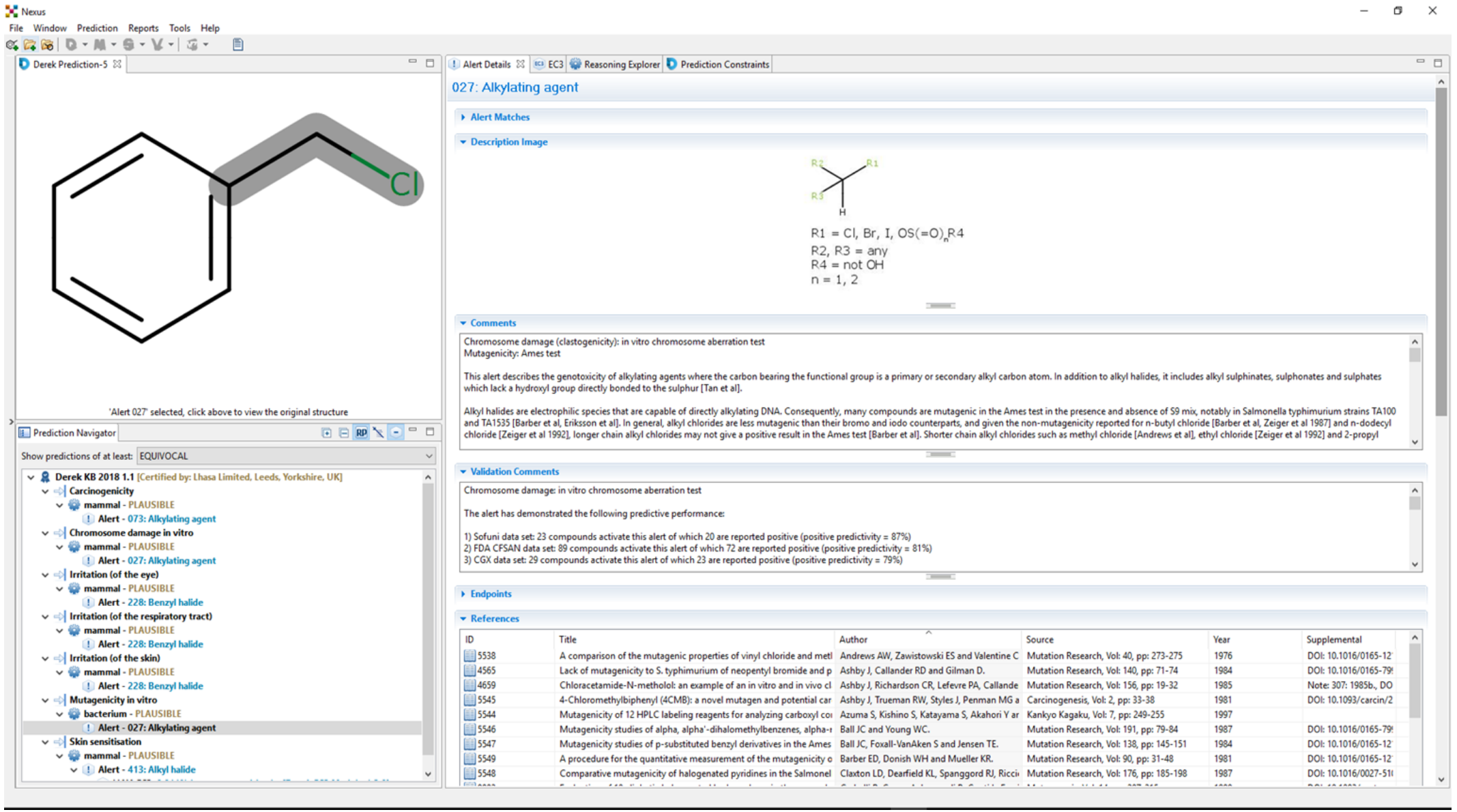


Alert fired

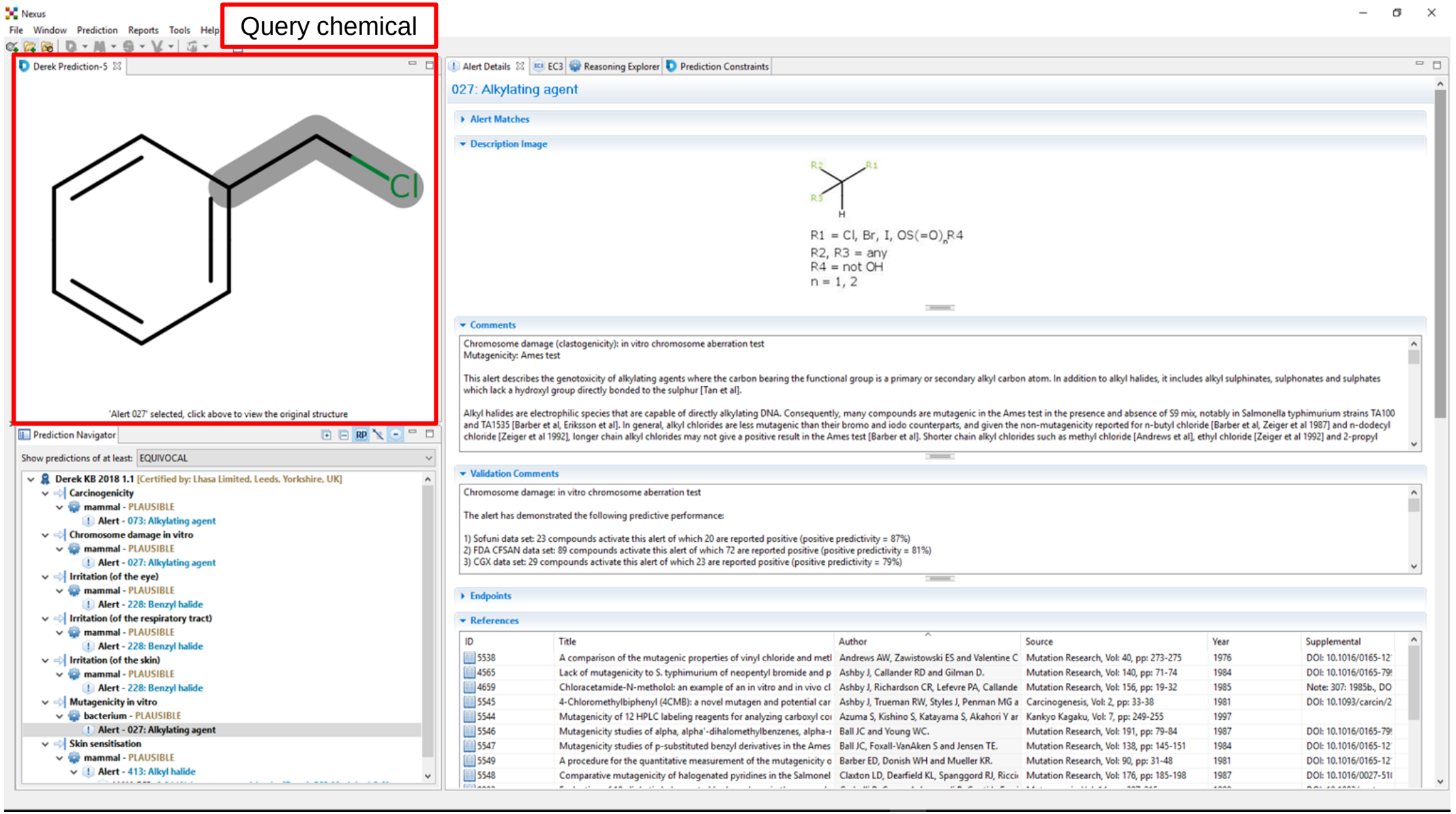


No alert fired

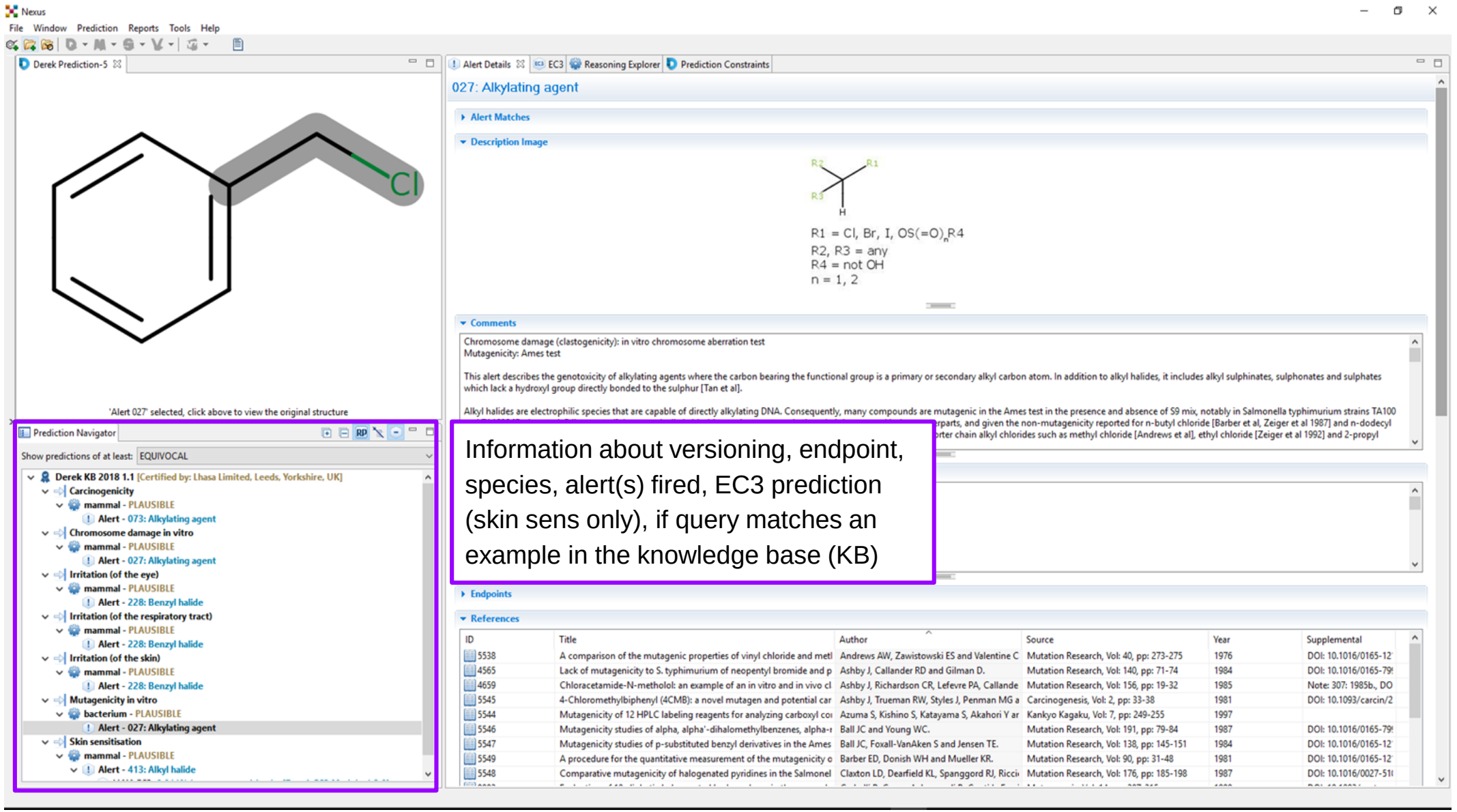
Example 1: Derek Nexus (Lhasa Limited)



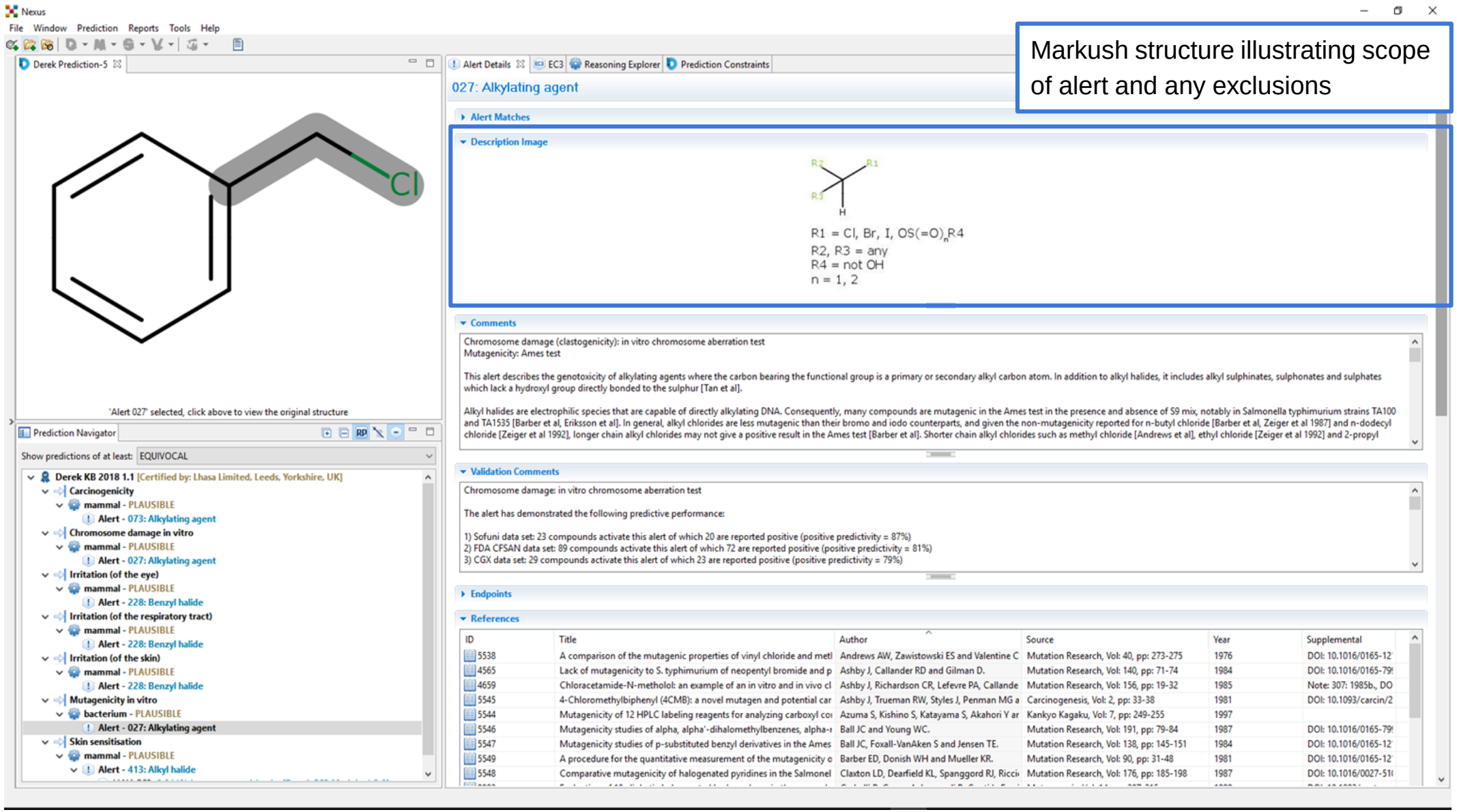
Example 1: Derek Nexus (Lhasa Limited)



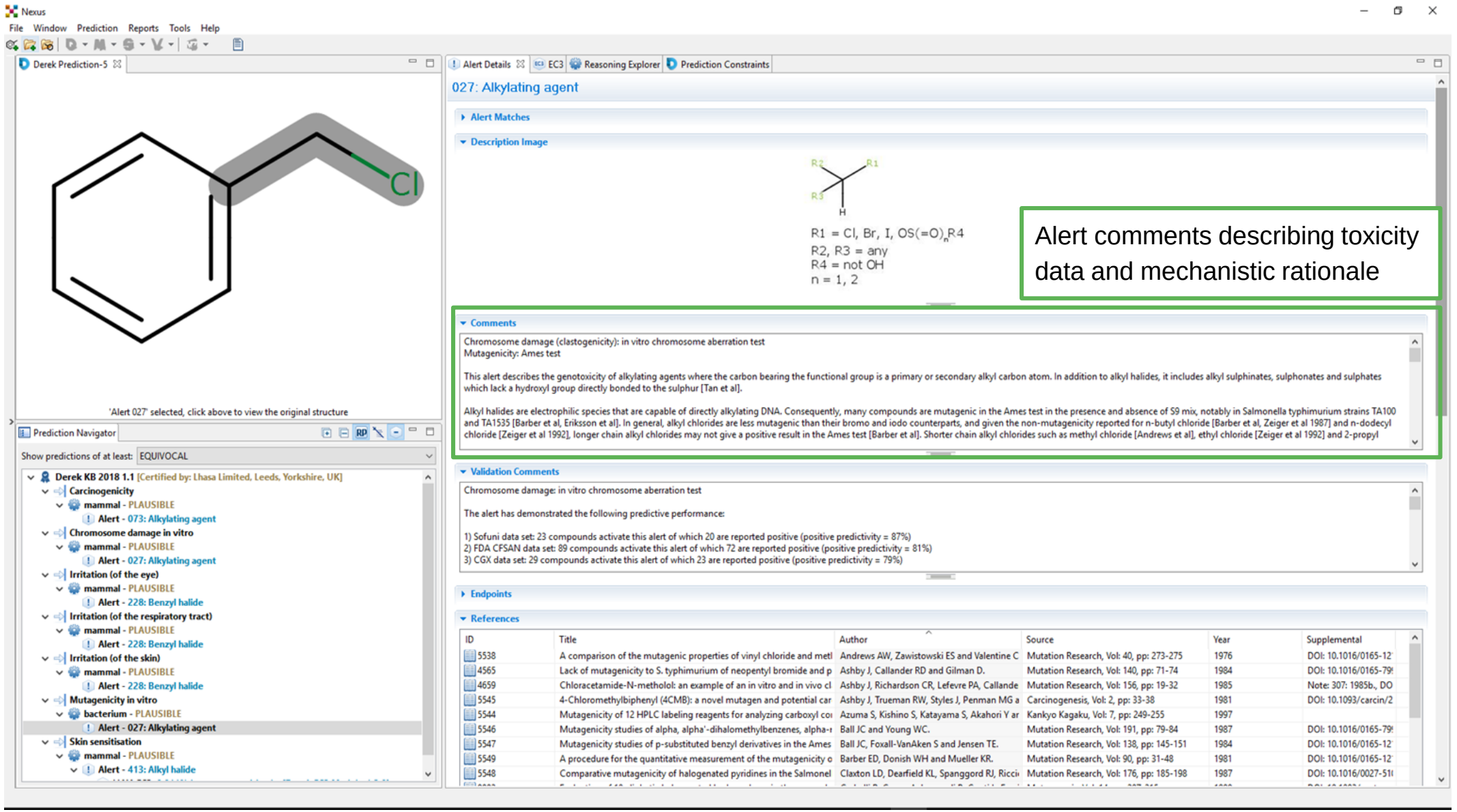
Example 1: Derek Nexus (Lhasa Limited)



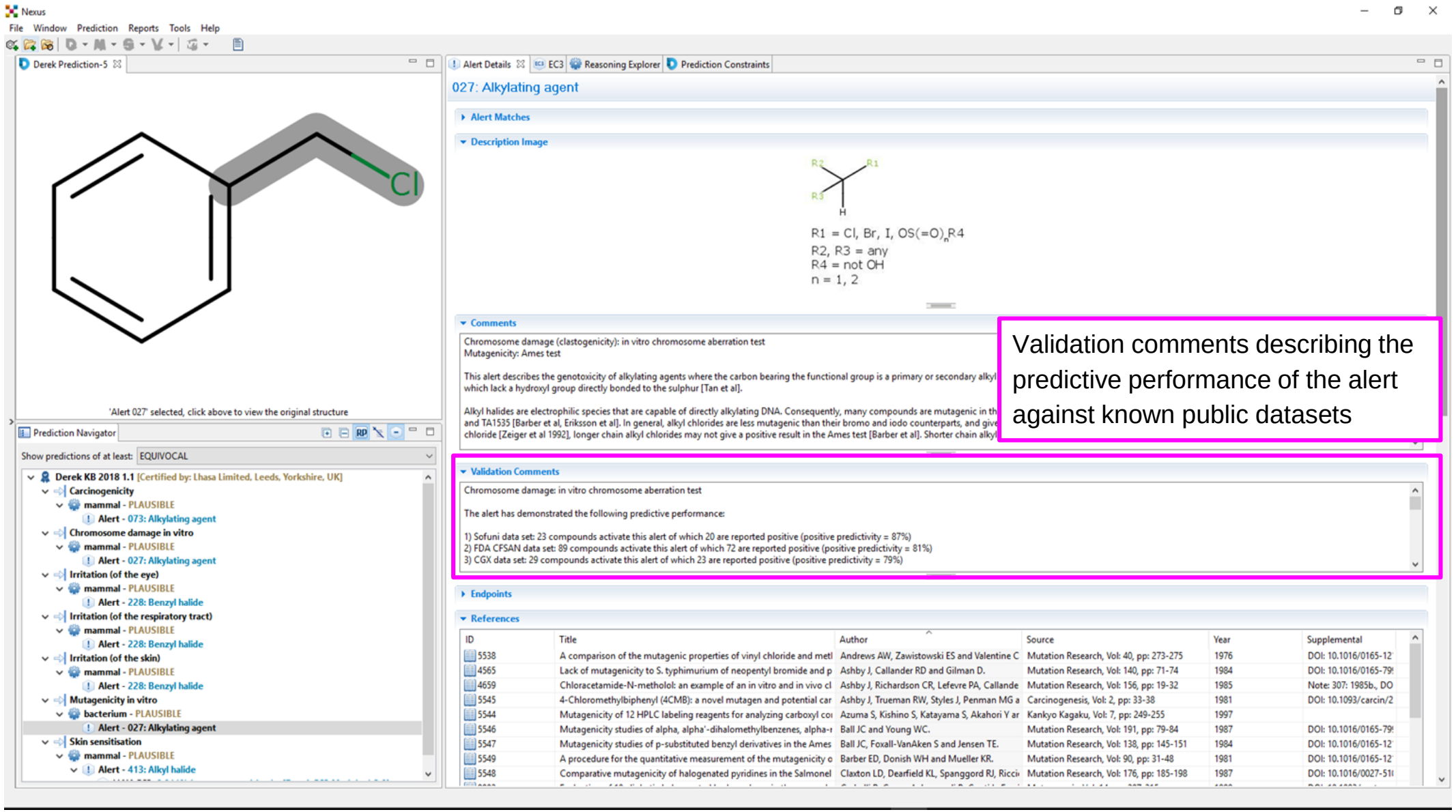
Example 1: Derek Nexus (Lhasa Limited)



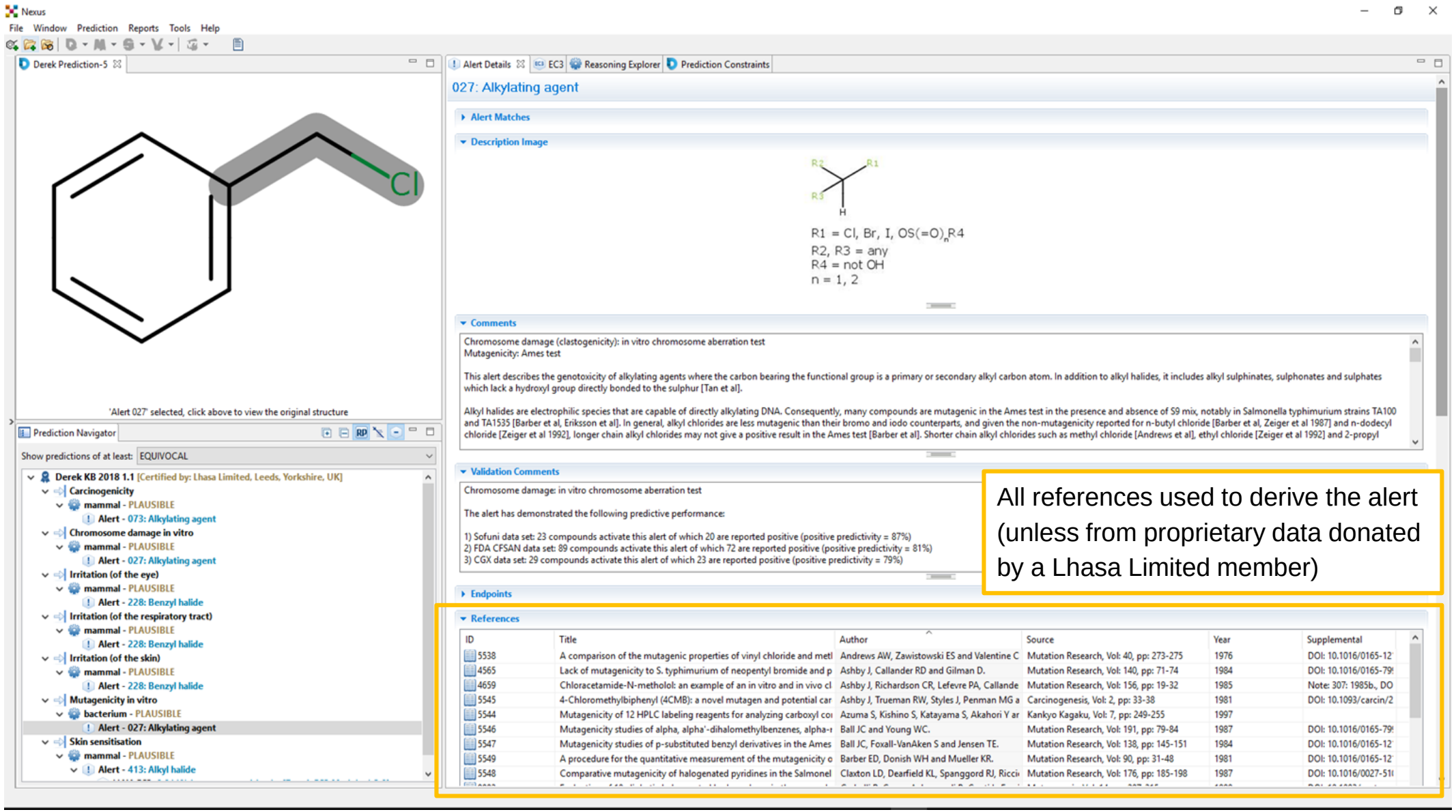
Example 1: Derek Nexus (Lhasa Limited)



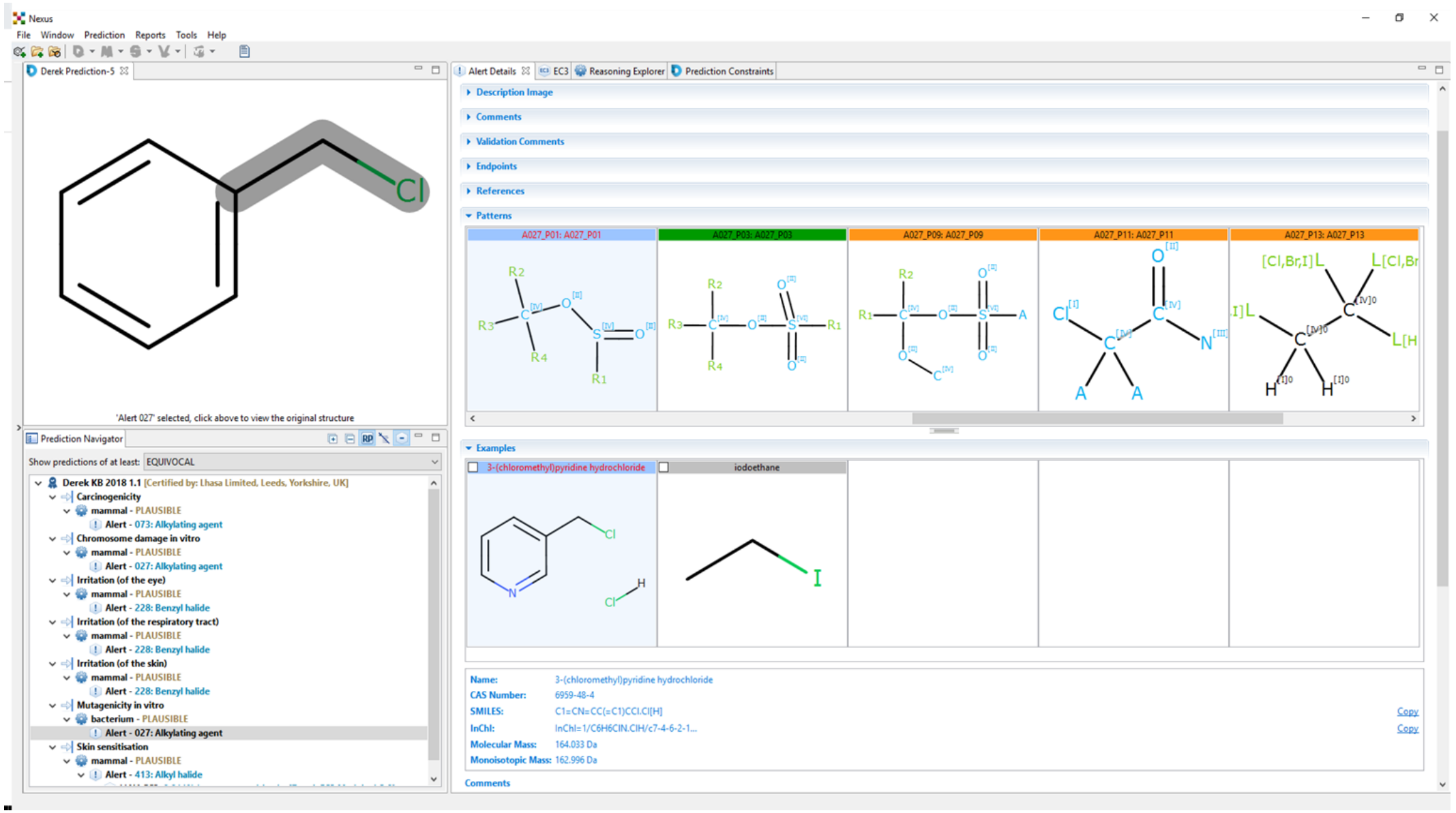
Example 1: Derek Nexus (Lhasa Limited)



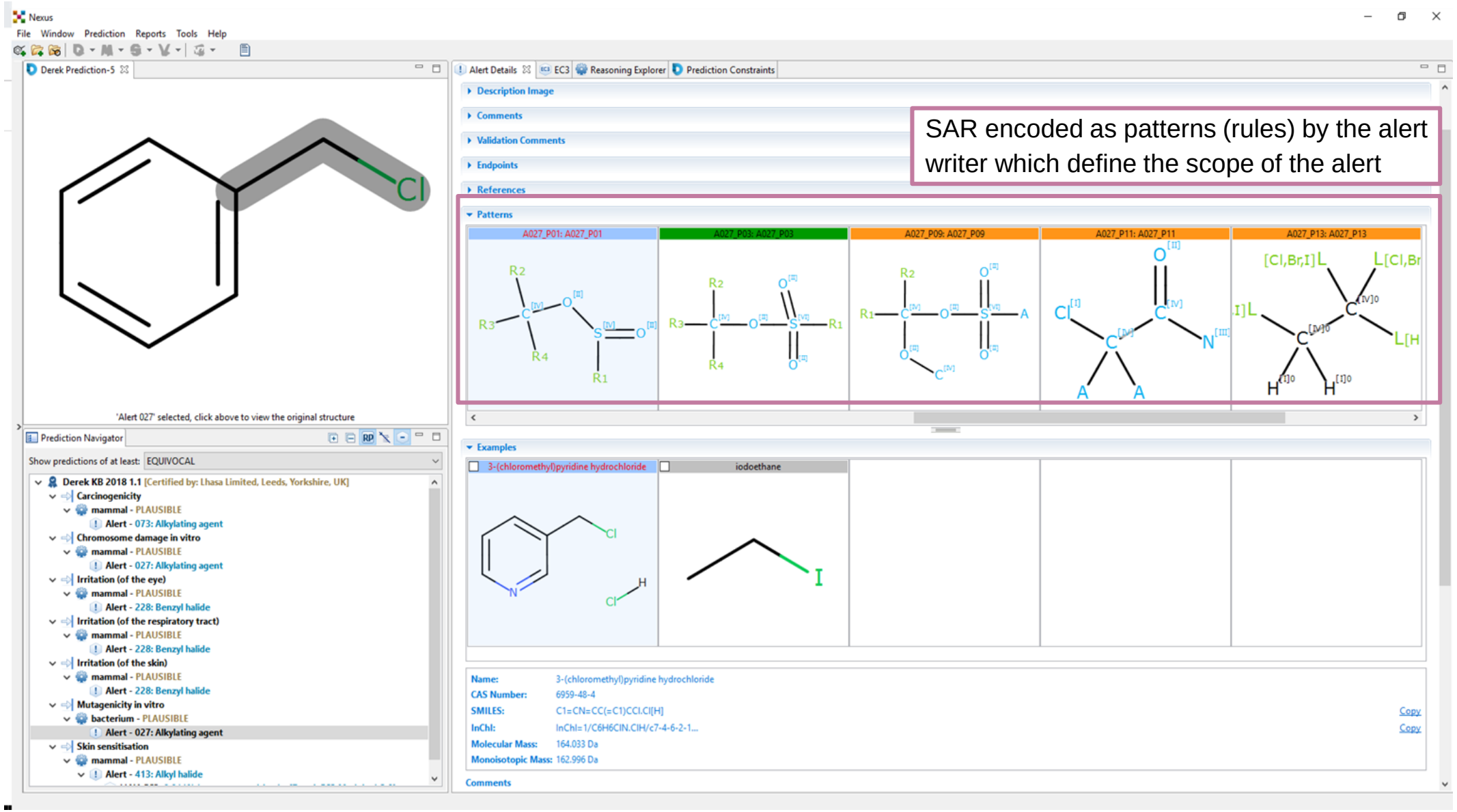
Example 1: Derek Nexus (Lhasa Limited)



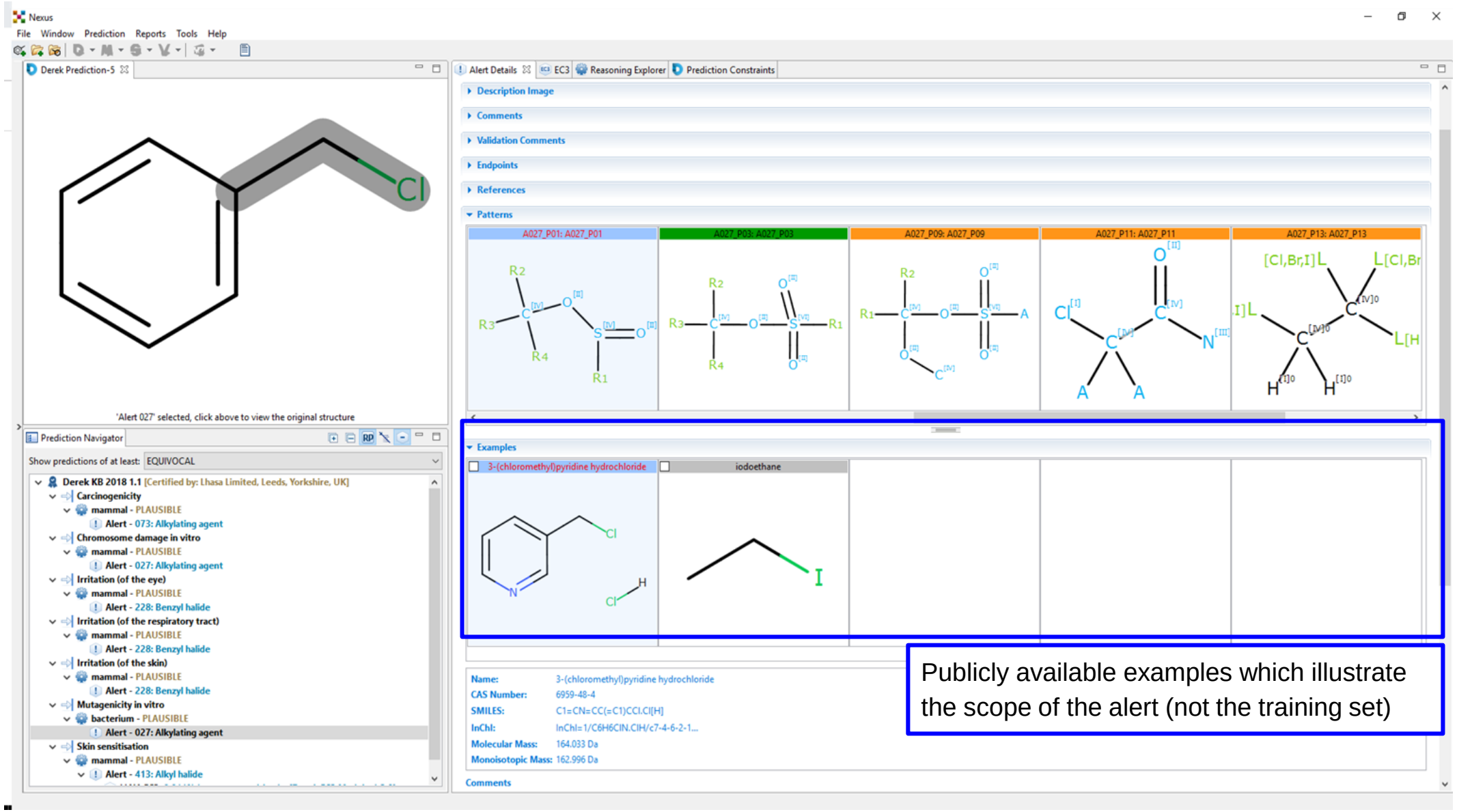
Example 1: Derek Nexus (Lhasa Limited)



Example 1: Derek Nexus (Lhasa Limited)

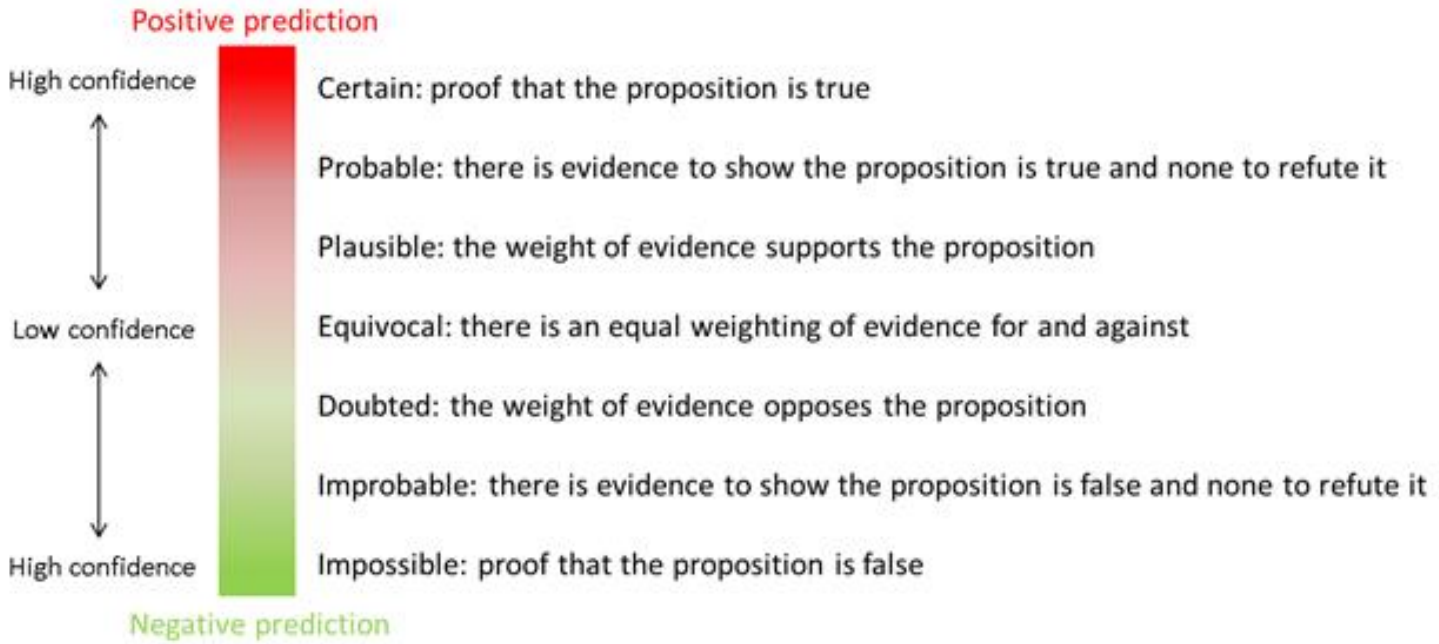


Example 1: Derek Nexus (Lhasa Limited)



Example 1: Derek Nexus (Lhasa Limited)

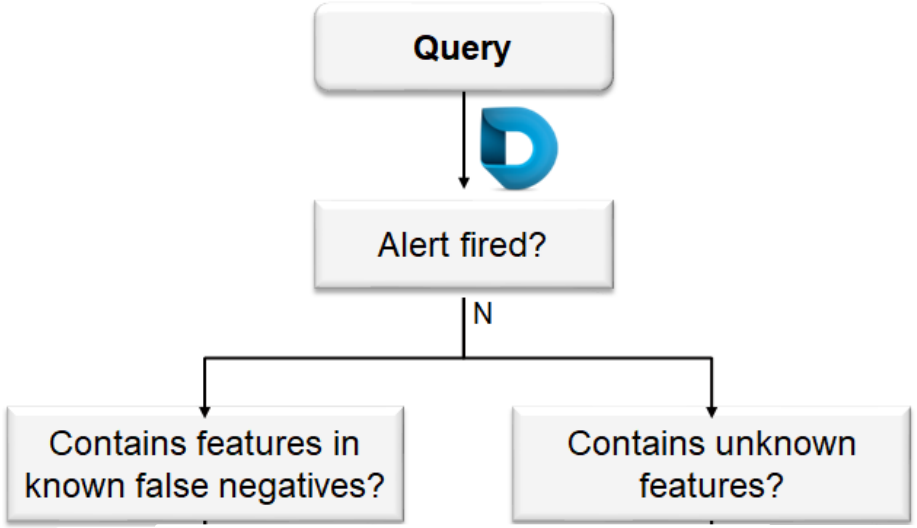
- When an alert is fired, the prediction is given as a **likelihood**
- Indicates **confidence** in prediction
  - Equivocal, plausible, probable, certain are all considered **positive** predictions
  - Doubted, improbable are very rarely observed
  - Impossible used when species chosen makes the adverse outcome unable to occur e.g. skin sensitisation prediction using bacterium as the species



Judson et al., *Toxicol. Res.*, 2013, 2, 70-79.

Example 1: Derek Nexus (Lhasa Limited)

- For *in vitro* mutagenicity and skin sensitisation a **negative prediction** model is available
- In the absence of an alert, a negative prediction is provided

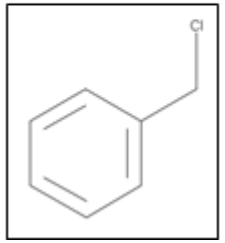


- **Misclassified** and **unclassified** features are **negative predictions** but highlight potentially concerning areas to consider reviewing to increase confidence

Williams et al., *Regul Toxicol Pharmacol*. 2016, 76, 79-86. Chilton et al., *Regul Toxicol Pharmacol*. 2018, 95, 227-235.

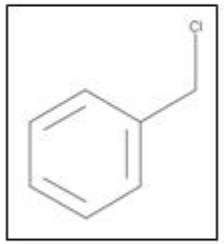
Example 2: Leadscope Model Applier

1. Enter Chemical



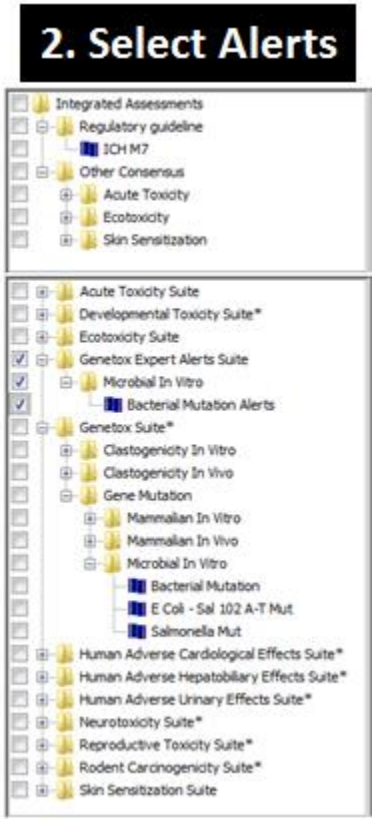
Example 2: Leadscope Model Applier

1. Enter Chemical



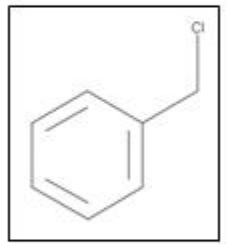
Read in chemical

2. Select Alerts



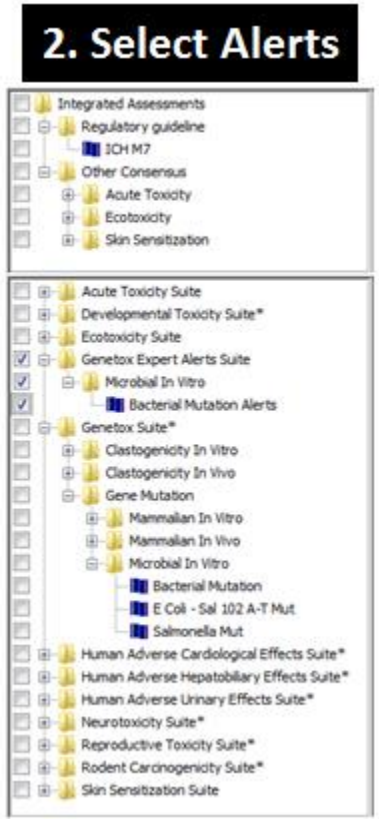
Example 2: Leadscope Model Applier

1. Enter Chemical



Read in chemical

2. Select Alerts



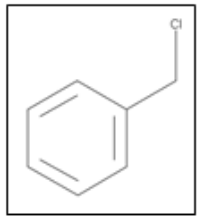
- Assess applicability domain
- Look-up any data
- Apply alerts

3. View prediction

Structure	Prediction	Precision	Matched Alerts
 SOT Webinar 1	Positive	0.8333	321: Aryl Methyl Halide (0.83)

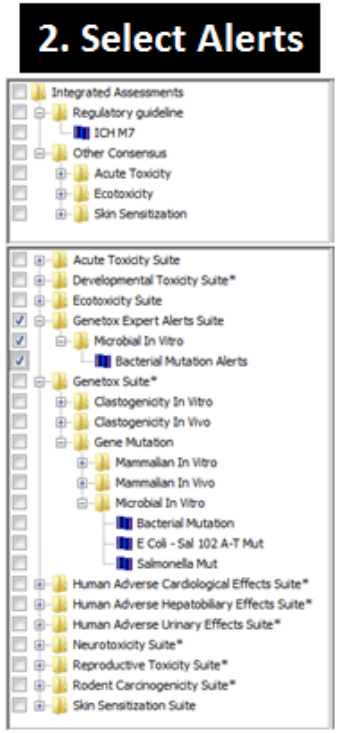
Example 2: Leadscope Model Applier

1. Enter Chemical



Read in chemical

2. Select Alerts



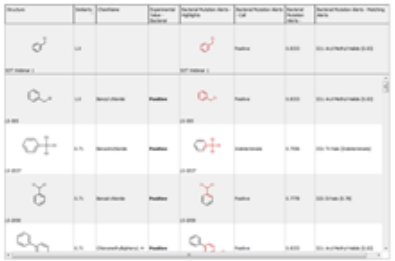
- Assess applicability domain
- Look-up any data
- Apply alerts

3. View prediction

Structure	Prediction	Precision	Matched Alerts
SOT Webinar 1	Positive	0.8333	321: Aryl Methyl Halide (0.83)

- Explain alert
- View analogs and underlying study report

4. Explain



- Analogs
- Experimental results
- Study report
- Prediction of analogs

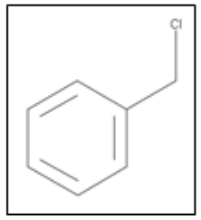
- Structural definition
- Deactivating Factors
- Sensitive tester strains
- Mechanisms
- Examples
- References

Example 2: Leadscope Model Applier

	Public		Proprietary								Total	
	Neg	Pos	Neg	Pos	Neg	Pos	Neg	Pos	Neg	Pos	Neg	Pos
852 321: Aryl Methyl Halide 1	10	34	[redacted]								59	106
853 321 - mitigating: Alkyl Methyl Halide Mitigator 1	1	0									4	0
854 321 - mitigating: Alkyl Methyl Halide Mitigator 2	0	0									4	0
855 321 - mitigating: Alkyl Methyl Halide Mitigator 3	0	0									3	0
856 321 - mitigating: Alkyl Methyl Halide Mitigator 4	1	0									5	1
857 321: Aryl Methyl Halide 2	0	6									6	27

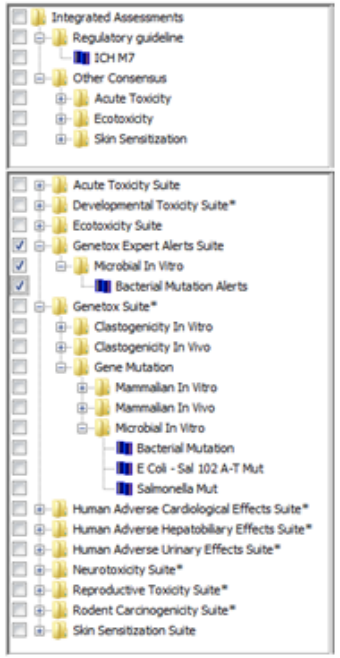
Qualify activating and deactivating features based on public and proprietary data

1. Enter Chemical



Read in chemical

2. Select Alerts



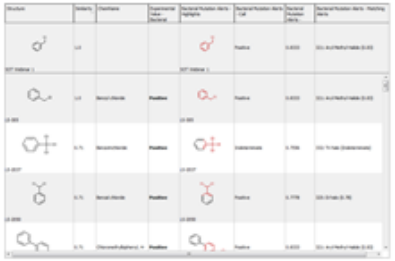
- Assess applicability domain
- Look-up any data
- Apply alerts

3. View prediction

Structure	Prediction	Precision	Matched Alerts
 SOT Webinar 1	Positive	0.8333	321: Aryl Methyl Halide (0.83)

- Explain alert
- View analogs and underlying study report

4. Explain

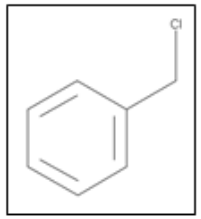


- Analogs
- Experimental results
- Study report
- Prediction of analogs

- Structural definition
- Deactivating Factors
- Sensitive tester strains
- Mechanisms
- Examples
- References

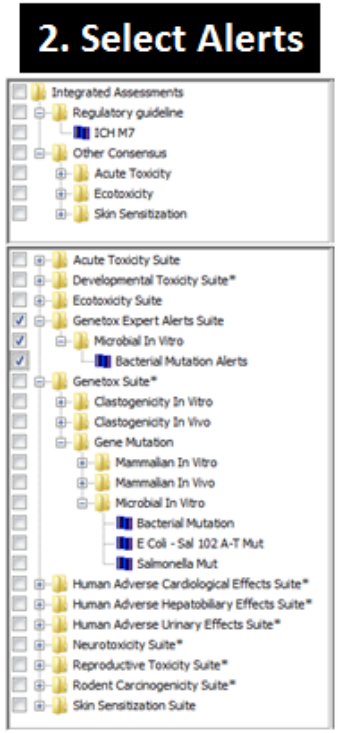
Example 2: Leadscope Model Applier

1. Enter Chemical



Read in chemical

2. Select Alerts



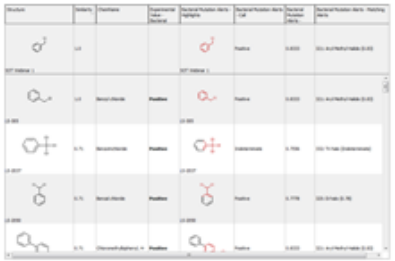
- Assess applicability domain
- Look-up any data
- Apply alerts

3. View prediction

Structure	Prediction	Precision	Matched Alerts
SOT Webinar 1	Positive	0.8333	321: Aryl Methyl Halide (0.83)

- Explain alert
- View analogs and underlying study report

4. Explain

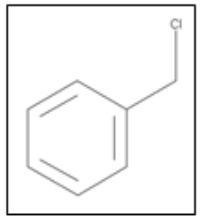


- Analogs
- Experimental results
- Study report
- Prediction of analogs

- Structural definition
- Deactivating Factors
- Sensitive tester strains
- Mechanisms
- Examples
- References

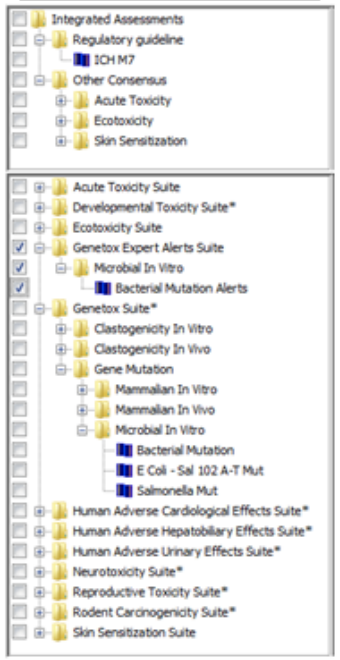
Example 2: Leadscope Model Applier

1. Enter Chemical



Read in chemical

2. Select Alerts



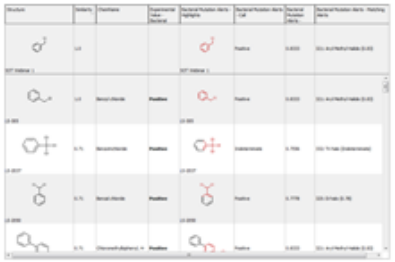
- Assess applicability domain
- Look-up any data
- Apply alerts

3. View prediction

Structure	Prediction	Precision	Matched Alerts
SOT Webinar 1	Positive	0.8333	321: Aryl Methyl Halide (0.83)

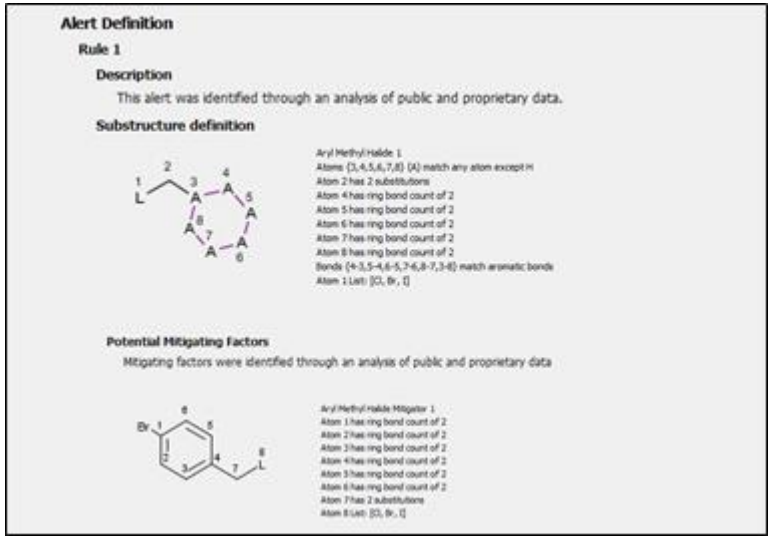
- Explain alert
- View analogs and underlying study report

4. Explain

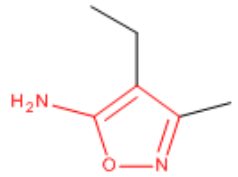


- Analogs
- Experimental results
- Study report
- Prediction of analogs

- Structural definition
- Deactivating Factors
- Sensitive tester strains
- Mechanisms
- Examples
- References

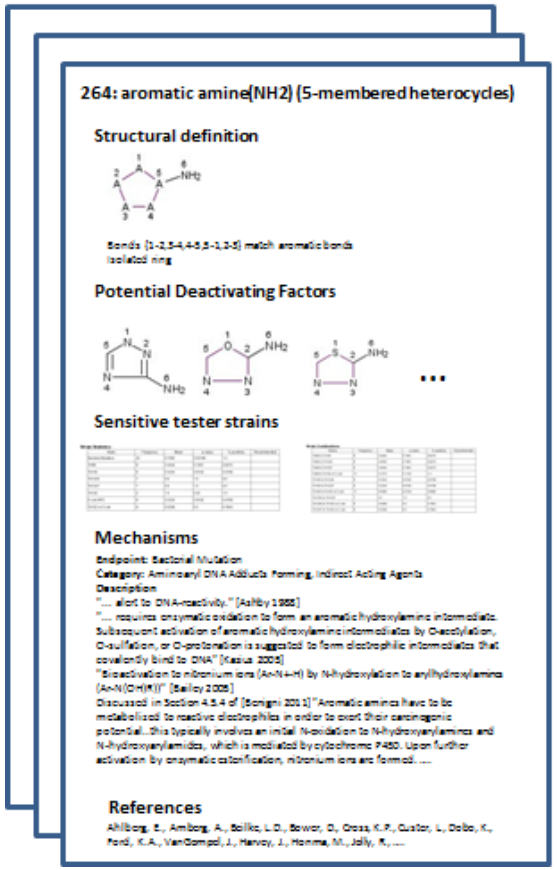


Example 2: Leadscope Model Applier

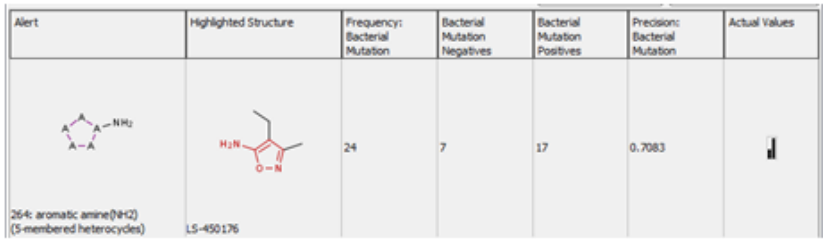


Prediction:	Mutagenic
Alert:	264: aromatic amine (NH2) (5-membered heterocycles)

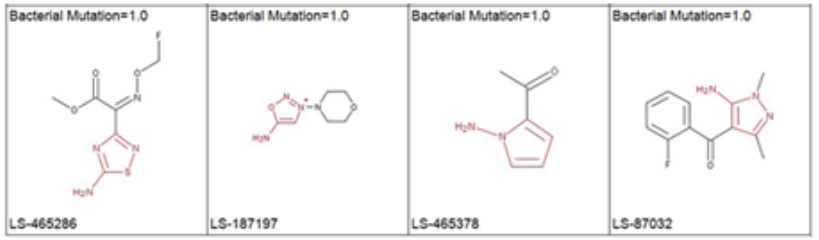
Alert definition



Justification



Matching examples



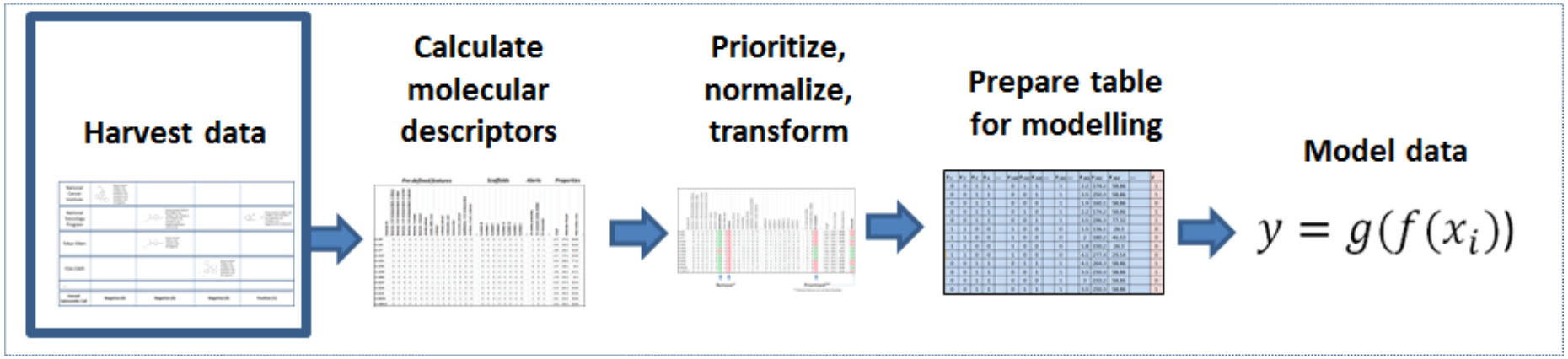
Statistical-based methodologies

Computational Toxicology
Specialty Section

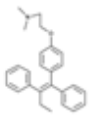
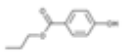
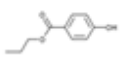
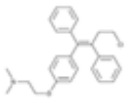
Statistical-based methodologies (QSAR)

- A training set is compiled with the chemical structure and the corresponding toxicity
- Molecular descriptors are calculated for each chemical structure
- A model is built using one or more molecular descriptors from every chemical to predict the toxic response
- Examples of statistical models are:
  - Random Forests
  - Artificial Neural Networks
  - Deep Learning
  - k-Nearest Neighbors
  - Support Vector Machines
  - Partial Logistic Regression
  - Partial Least Squares Regression
  - Multiple Linear Regression

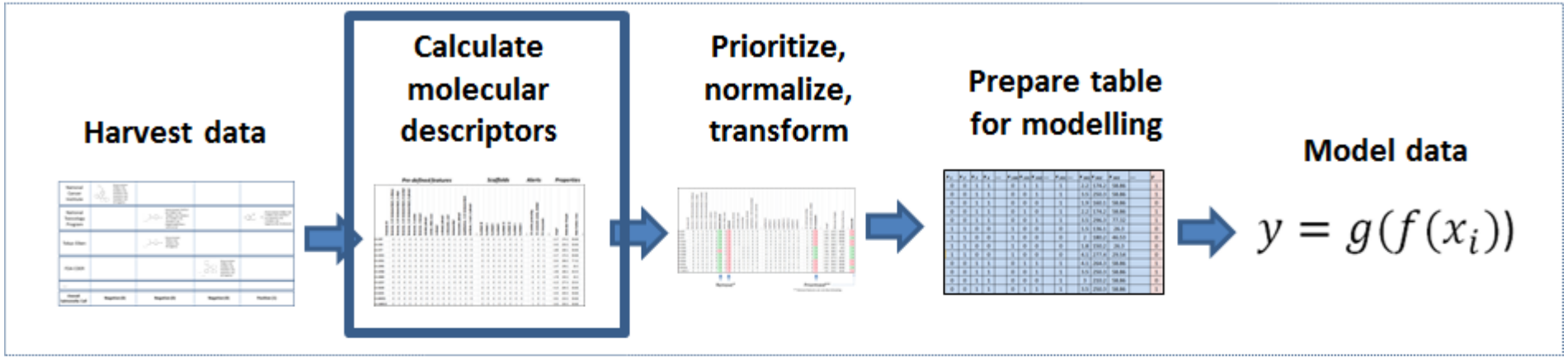
Building statistical-based models

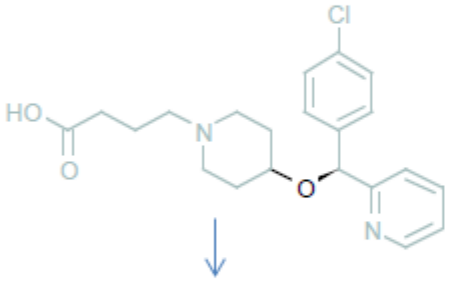


Harvest data

National Cancer Institute	 Strains tested: TA98(+/-s9), TA100(+/-s9), TA1535(+/-s9), TA1537(+/-s9), TA1538(+/-s9) All negative			
National Toxicology Program		 Strains tested: TA97(+/-s9), TA98(+/-s9), TA100(+/-s9), TA102(+/-s9), TA104(+/-s9), TA1535(+/-s9) Negative(40)/Not a valid test (4)		 Strains tested: TA98(+/-s9), TA100(+/-s9), TA1535(+/-s9), TA1537(+/-s9), TA1538(+/-s9) Negative (78), Positive (12)
Tokyo Eiken		 Strains tested: TA97(+/-s9), TA102(+/-s9), All negative		
FDA CDER			 Strains tested: TA98(+/-s9), TA100(+/-s9), TA1535(+/-s9), TA1537(+/-s9) All negative	
....				
Overall Call	**Negative (0)**	**Negative (0)**	**Negative (0)**	**Positive (1)**

Building statistical-based models



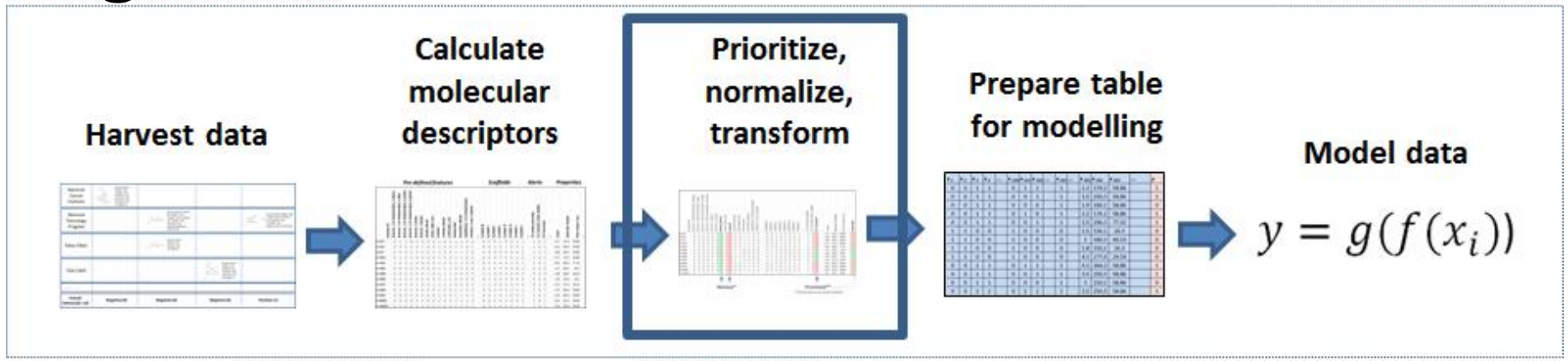


X_1: carbonyl, alkyl, acyc-
X_2: ALogP = 4.17
X_3: tert-amine
X_4: Hydrogen Bond Donors = 0
X_5: Parent Atom Count = 27
X_6: Parent Molecular Weight = 388.9
X_7: Hydrogen Bond Acceptors = 4
X_8: piperidine
X_9: Polar Surface Area = 62.66
X_{10}: Rotatable Bonds = 8
X_{11}: propane
X_{12}: carboxylic acid, alkyl-
X_{13}: benzene, 1-chloro (isolated)
X_{14}: butane
X_{15}: carboxy
X_{16}: Scaffold 5002
X_{17}: piperidine, 1-(alkyl, acyc)-

Calculate molecular descriptors

Structure ID	benzene, 1-(N-iminomethyl)-,2-alkoxy-	benzene, 1-(N-iminomethyl)-,2-alkyl-	benzene, 1-(N-iminomethyl)-,4-methyl-	benzene, 1-(N-iminomethyl)-,4-phenyl-	benzene, 1-amino-	benzene, 1-benzyl-	alkohol, phenyl-	amine, alkyl, acyc-	carbonyl	carbonyl, phenyl-	carboxylate, aryl-	iminomethyl	isocyanate, phenyl-	naphthalene, 1-(N-iminomethyl)-	methane, 1-aryl-,1-phenyl-	...	Scaffold 10	Scaffold 3	Scaffold 9	Scaffold 8	Scaffold 11	Scaffold 12	Scaffold 5	Scaffold 6	...	23: amine generating	25: aromatic amine, dialkyl	93: isocyanate	...	ALogP	Molecular Weight	Polar Surface Area	...
LS-247	0	1	1	0	0	0	0	0	1	0	0	1	1	0	0		0	0	0	1	0	0	1	1		1	0	1		2.17	174.2	58.86	
LS-334	0	0	0	0	0	1	0	0	1	0	0	1	1	0	1		0	0	1	0	1	0	0	1		1	0	1		3.45	250.3	58.86	
LS-377	0	0	0	0	0	0	0	0	1	0	0	1	1	0	0		0	0	0	0	0	0	0	0		1	0	1		1.86	160.1	58.86	
LS-1915	0	1	0	0	0	0	0	0	1	0	0	1	1	0	0		0	0	0	1	0	0	1	0		1	0	1		2.17	174.2	58.86	
LS-1931	1	0	0	1	0	0	0	0	1	0	0	1	1	0	0		0	0	0	0	0	1	0	1		1	0	1		3.54	296.3	77.32	
LS-1945	0	0	0	0	0	0	0	0	1	1	1	0	0	0	0		0	1	0	0	0	0	0	0		0	0	0		1.47	136.1	26.3	
LS-1958	0	0	0	0	0	0	1	0	1	1	1	0	0	0	0		1	1	0	0	0	0	0	0		0	0	0		1.96	180.2	46.53	
LS-2069	0	0	0	0	0	0	0	0	1	1	1	0	0	0	0		0	1	0	0	0	0	0	0		0	0	0		1.78	150.2	26.3	
LS-2257	0	0	0	0	1	0	0	1	1	1	1	0	0	0	0		1	1	0	0	0	0	0	0		0	1	0		4.13	277.4	29.54	
LS-5630	0	1	0	1	0	0	0	0	1	0	0	1	1	0	0		0	0	0	0	0	1	1	1		1	0	1		4.14	264.3	58.86	
LS-6191	0	0	0	0	0	1	0	0	1	0	0	1	1	0	1		0	0	1	0	1	0	0	1		1	0	1		3.45	250.3	58.86	
LS-84455	0	0	0	0	0	0	0	0	1	0	0	1	1	1	0		0	0	0	0	0	0	0	0		1	0	1		3.01	210.2	58.86	
LS-168513	0	1	0	0	0	1	0	0	1	0	0	1	1	0	1		0	0	1	0	0	0	1	1		1	0	1		3.45	250.3	58.86	

Building statistical-based models



Prioritize, normalize, transform

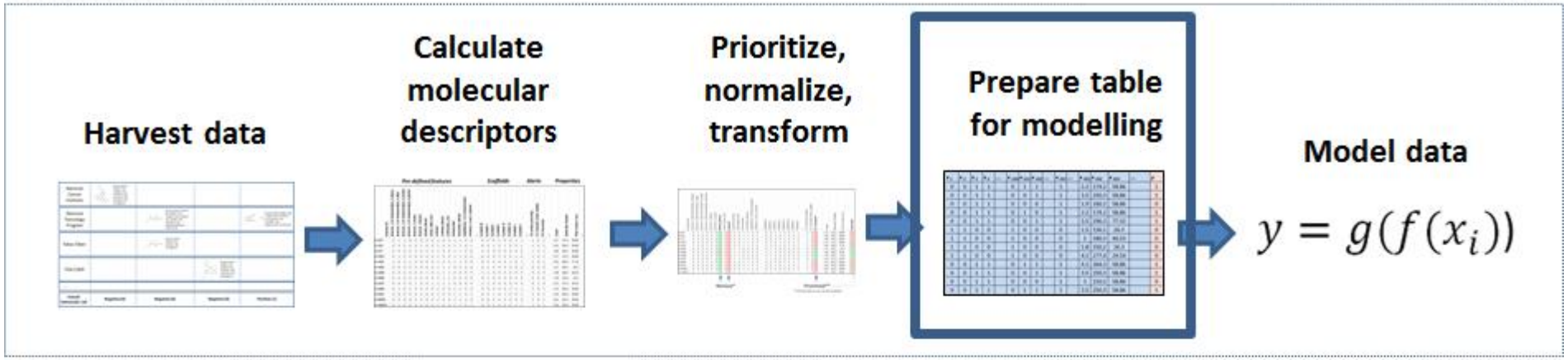
Structure ID	benzene, 1-(bi-iminoarethyl)-,2-alkoxy-	benzene, 1-(bi-iminoarethyl)-,2-alkyl-	benzene, 1-(bi-iminoarethyl)-,4-methyl-	benzene, 1-(bi-iminoarethyl)-,4-phenyl-	benzene, 1-amino-	benzene, 1-benzyl-	**alcohol, phenyl**	amine, alkyl, acyc-	**carbonyl**	carbonyl, phenyl-	carboxylate, aryl-	iminomethyl	isocyanate, phenyl-	naphthalene, 1-(bi-iminoarethyl)-	methane, 1-aryl-,1-phenyl-	...	Scaffold 10	Scaffold 3	Scaffold 9	Scaffold 8	Scaffold 11	Scaffold 12	Scaffold 5	Scaffold 6	...	23: amine generating	25: aromatic amine, dialkyl	**93: isocyanate**	...	ALogP	Molecular Weight	Polar Surface Area	...	**Salmonella**
LS-247	0	1	1	0	0	0	0	0	1	0	0	1	1	0	0		0	0	0	1	0	0	1	1		1	0	1		2.17	174.2	58.86		1
LS-334	0	0	0	0	0	1	0	0	1	0	0	1	1	0	1		0	0	1	0	1	0	0	1		1	0	1		3.45	250.3	58.86		1
LS-377	0	0	0	0	0	0	0	0	1	0	0	1	1	0	0		0	0	0	0	0	0	0	0		1	0	1		1.86	160.1	58.86		0
LS-1915	0	1	0	0	0	0	0	0	1	0	0	1	1	0	0		0	0	0	1	0	0	1	0		1	0	1		2.17	174.2	58.86		1
LS-1931	1	0	0	1	0	0	0	0	1	0	0	1	1	0	0		0	0	0	0	0	1	0	1		1	0	1		3.54	296.3	77.32		1
LS-1945	0	0	0	0	0	0	0	0	1	1	1	0	0	0	0		0	1	0	0	0	0	0	0		0	0	0		1.47	136.1	26.3		0
LS-1958	0	0	0	0	0	0	1	0	1	1	1	0	0	0	0		1	1	0	0	0	0	0	0		0	0	0		1.96	180.2	46.53		0
LS-2069	0	0	0	0	0	0	0	0	1	1	1	0	0	0	0		0	1	0	0	0	0	0	0		0	0	0		1.78	150.2	26.3		0
LS-2257	0	0	0	0	1	0	0	1	1	1	1	0	0	0	0		1	1	0	0	0	0	0	0		0	1	0		4.13	277.4	29.54		0
LS-5630	0	1	0	1	0	0	0	0	1	0	0	1	1	0	0		0	0	0	0	0	1	1	1		1	0	1		4.14	264.3	58.86		1
LS-6191	0	0	0	0	0	1	0	0	1	0	0	1	1	0	1		0	0	1	0	1	0	0	1		1	0	1		3.45	250.3	58.86		1
LS-84455	0	0	0	0	0	0	0	0	1	0	0	1	1	1	0		0	0	0	0	0	0	0	0		1	0	1		3.01	210.2	58.86		0
LS-168513	0	1	0	0	0	1	0	0	1	0	0	1	1	0	1		0	0	1	0	0	0	1	1		1	0	1		3.45	250.3	58.86		1

Remove*

Prioritized**

** Remove features are not discriminating

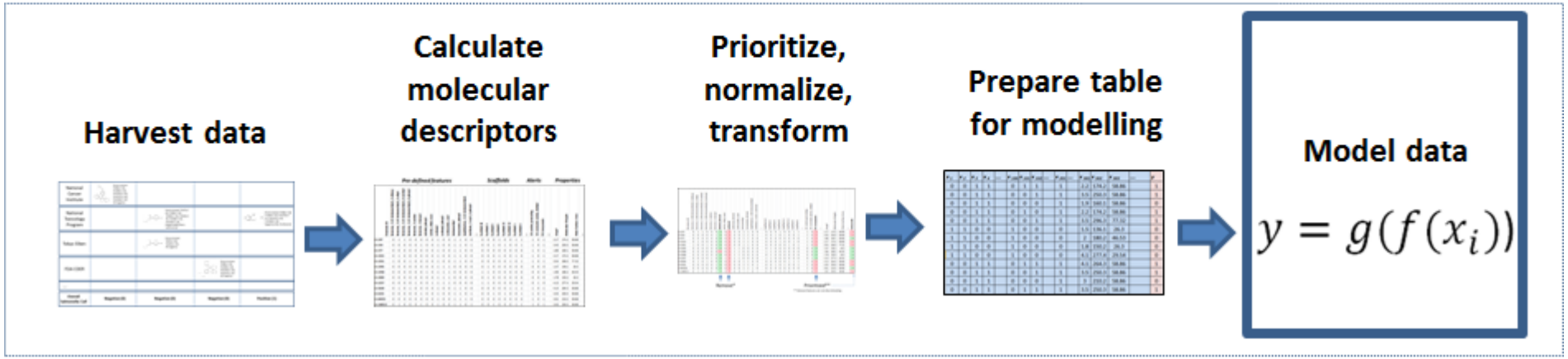
Building statistical-based models



Prepare table for modelling

x_1	x_2	x_3	x_4	....	x_{100}	x_{101}	x_{102}	....	x_{201}	....	x_{301}	x_{302}	x_{303}	....	y
0	0	1	1		0	1	1		1		2.2	174.2	58.86		1
0	0	1	1		0	0	1		1		3.5	250.3	58.86		1
0	0	1	1		0	0	0		1		1.9	160.1	58.86		0
0	0	1	1		0	1	0		1		2.2	174.2	58.86		1
0	0	1	1		0	0	1		1		3.5	296.3	77.32		1
1	1	0	0		1	0	0		0		1.5	136.1	26.3		0
1	1	0	0		1	0	0		0		2	180.2	46.53		0
1	1	0	0		1	0	0		0		1.8	150.2	26.3		0
1	1	0	0		1	0	0		0		4.1	277.4	29.54		0
0	0	1	1		0	1	1		1		4.1	264.3	58.86		1
0	0	1	1		0	0	1		1		3.5	250.3	58.86		1
0	0	1	1		0	0	0		1		3	210.2	58.86		0
0	0	1	1		0	1	1		1		3.5	250.3	58.86		1

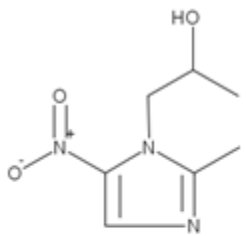
Building statistical-based models



Harvest data

$$y = g(f(x_i))$$

Making a prediction using statistical-based models



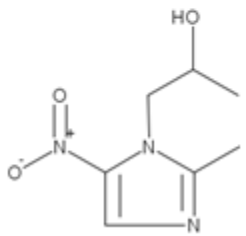
?

Applicability domain
e.g., in-domain

Prediction
e.g., mutagenic

Confidence
e.g., probability = 0.73

Making a prediction using statistical-based models



X_1: Aromatic nitro
X_2: Nitro
X_3: Diazole
X_4: Polar surface area = 83.87
X_5: Rotatable bonds = 3
X_6: Hydrogen bond donors =1
X_7: Molecular weight = 185.2
X_8: Atom count = 13
X_9: aLogP = 0.48
X_{10}: Ethane, 1-amino-, 2-hydroxy
X_{11}: Propane

$$y = g(f(x_i))$$

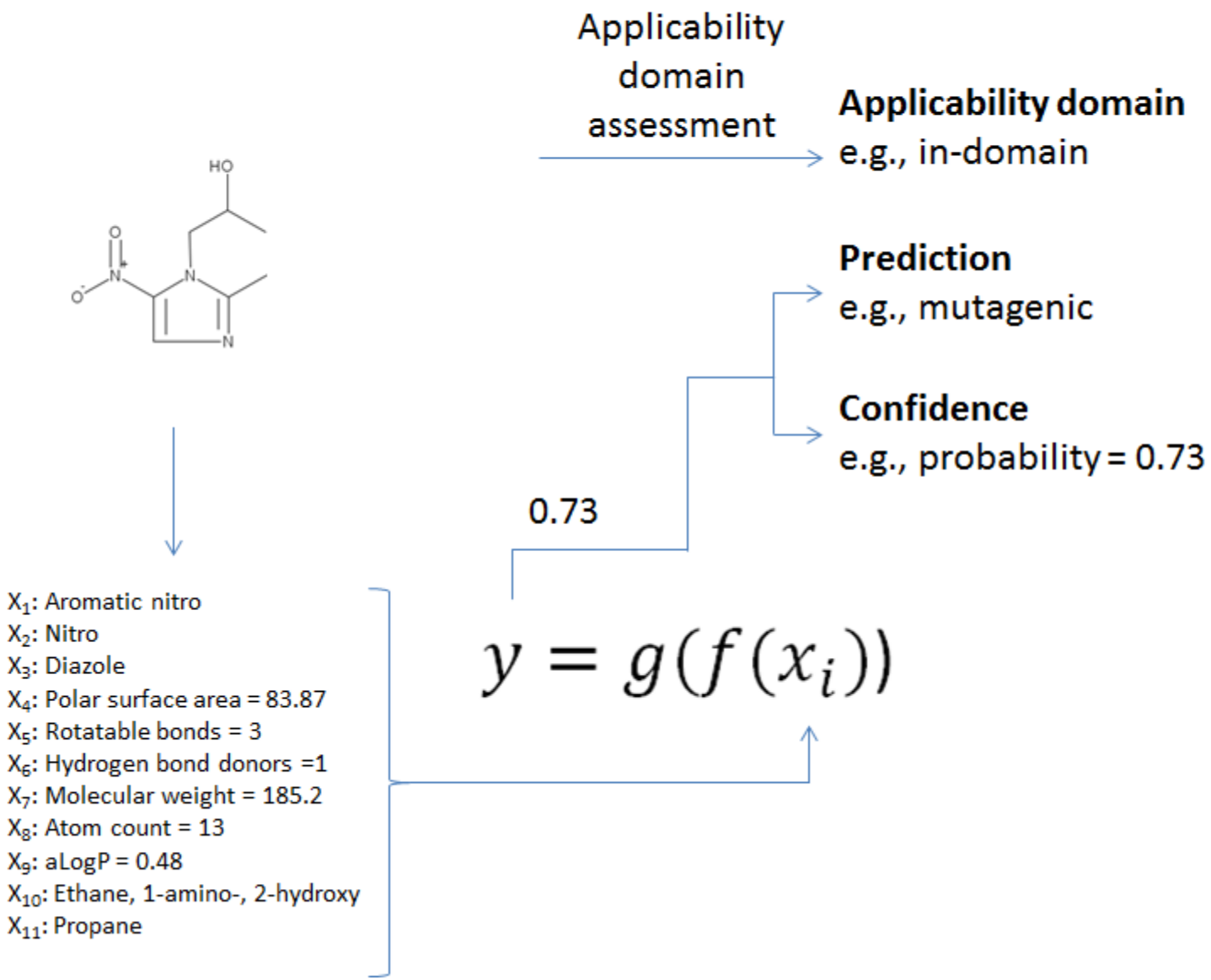
0.73

Applicability domain
e.g., in-domain

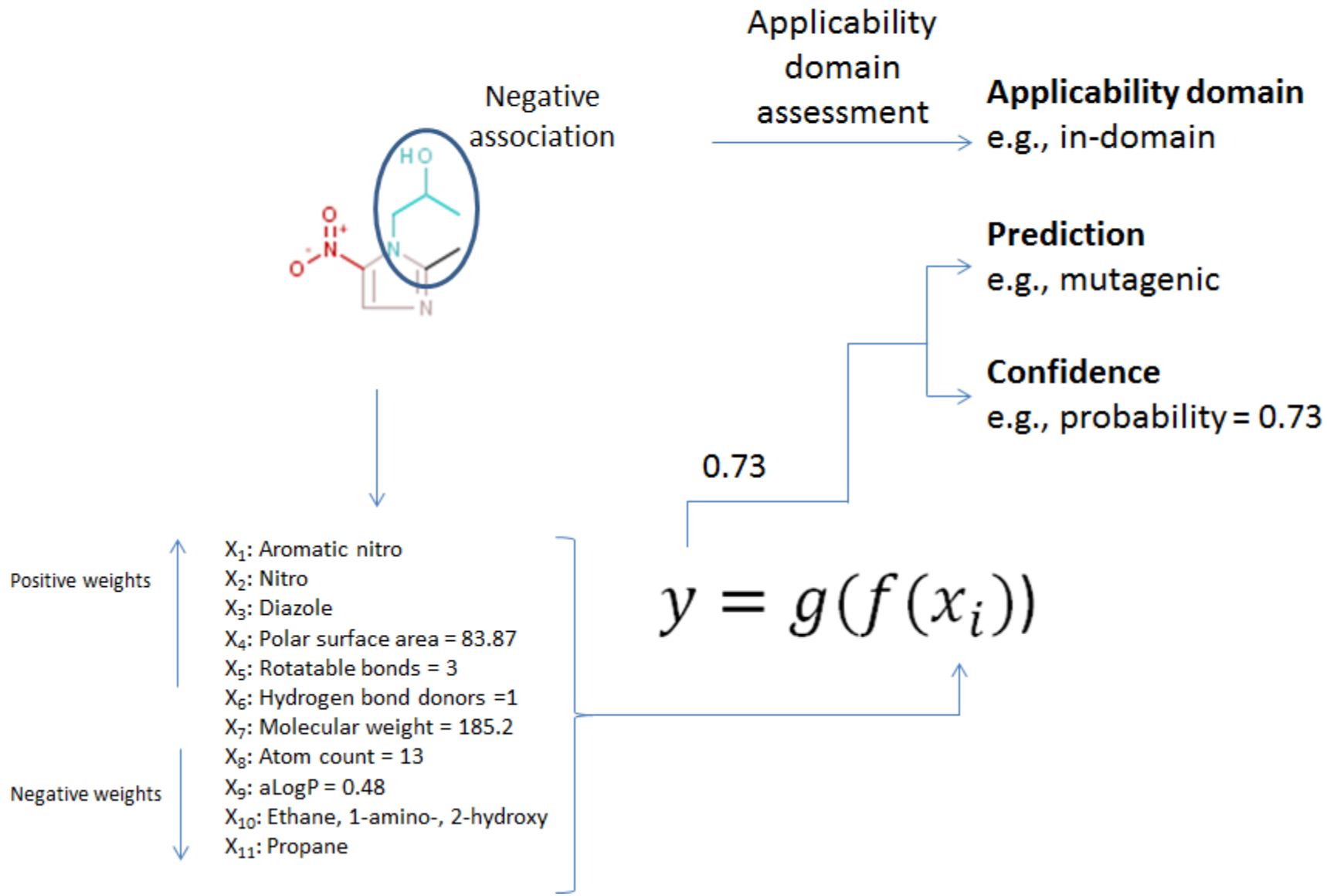
Prediction
e.g., mutagenic

Confidence
e.g., probability = 0.73

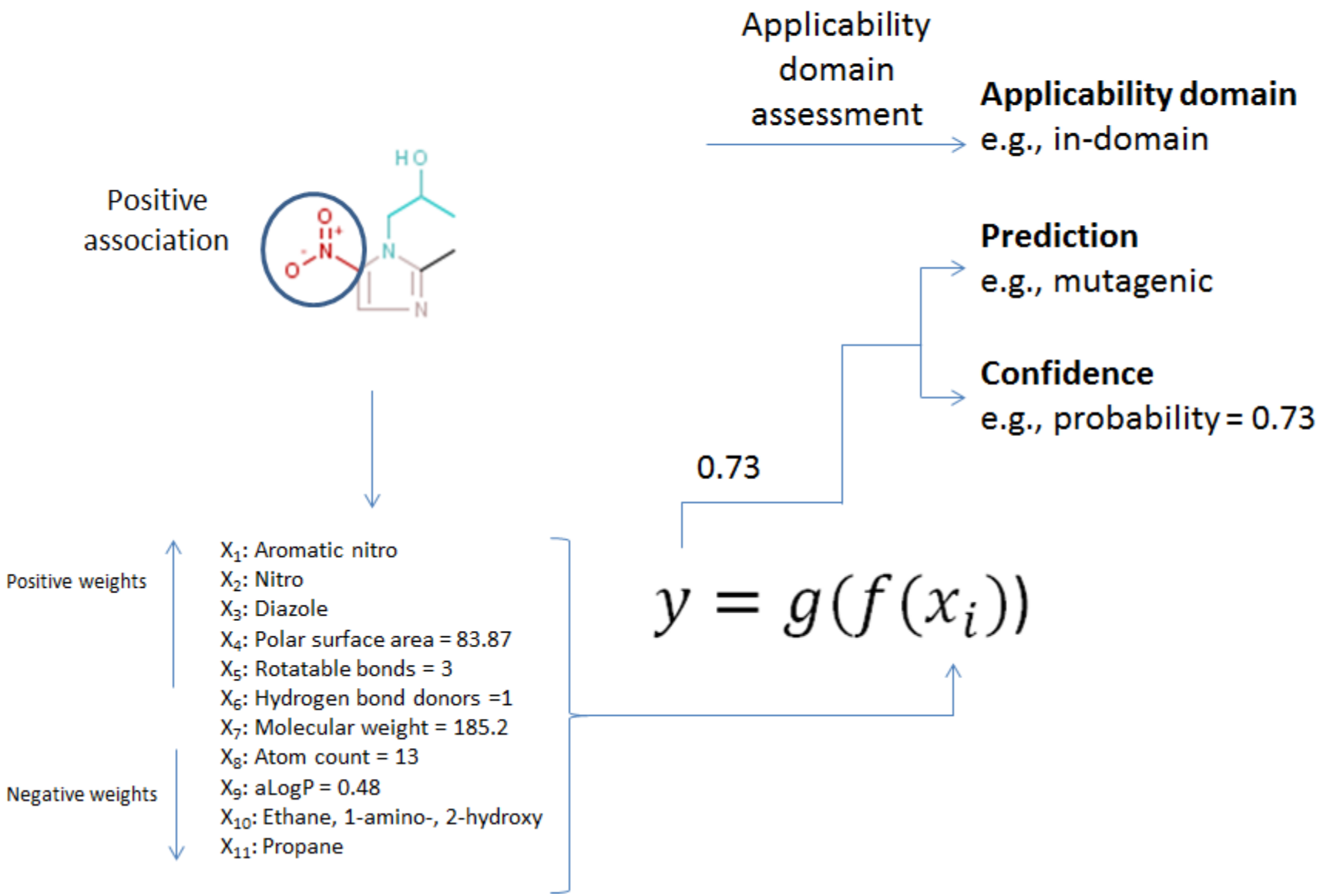
Making a prediction using statistical-based models



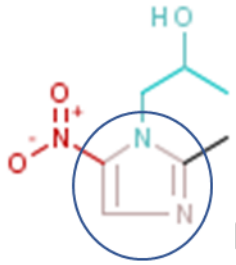
Making a prediction using statistical-based models



Making a prediction using statistical-based models



Making a prediction using statistical-based models



No significant positive or negative association

Positive weights

Negative weights

X_1: Aromatic nitro
X_2: Nitro
X_3: Diazole
X_4: Polar surface area = 83.87
X_5: Rotatable bonds = 3
X_6: Hydrogen bond donors =1
X_7: Molecular weight = 185.2
X_8: Atom count = 13
X_9: aLogP = 0.48
X_{10}: Ethane, 1-amino-, 2-hydroxy
X_{11}: Propane

$$y = g(f(x_i))$$

0.73

Applicability domain assessment

Applicability domain
e.g., in-domain

Prediction
e.g., mutagenic

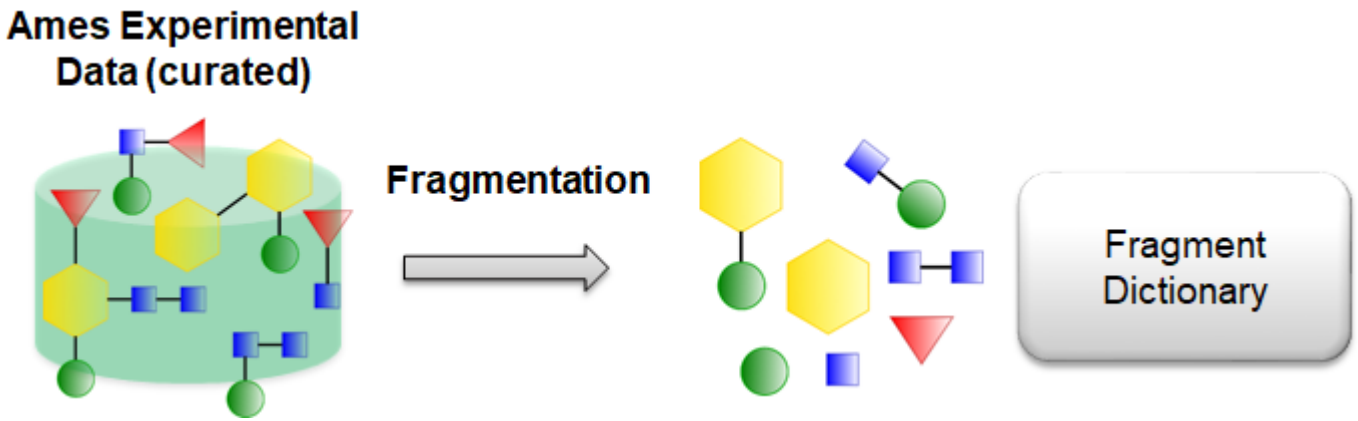
Confidence
e.g., probability = 0.73

Predicting mutagenicity using Sarah Nexus

- Statistical model for prediction of *in vitro* mutagenicity
- Model is built with Lhasa-curated public data and member-donated data
  - Ames Data, 5 strain, where available
  - 9882 unique structures (4716 positive / 5166 negative)

How Sarah works...

- DNA-reactivity is driven by chemical reactivity and is best modelled at the fragment-level
  - Molecules are fragmented according to a series of rules
- The Ames Test activity of the parent molecule of each fragment is retrieved

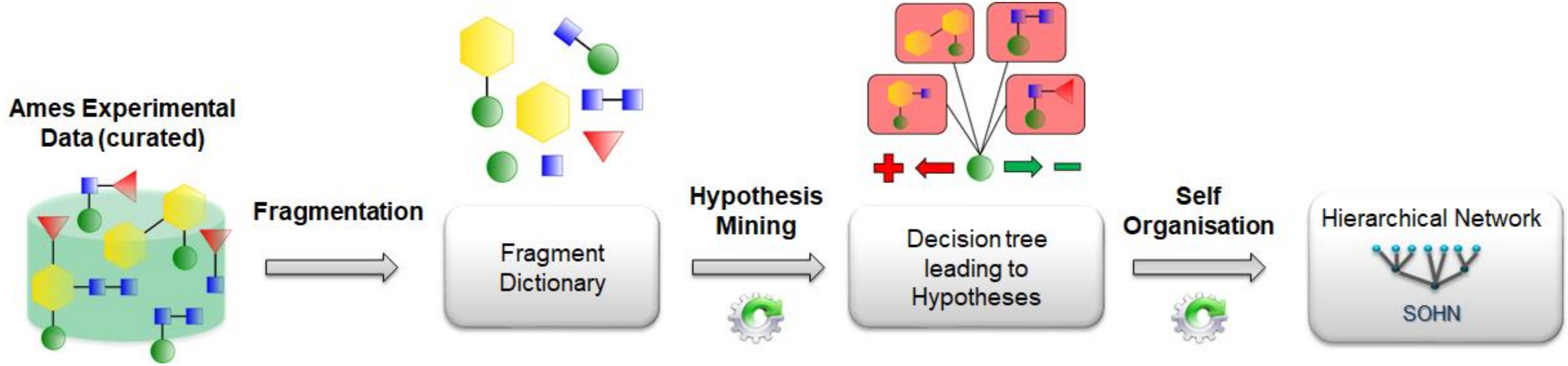


Hanser et al., *Journal of Cheminformatics*, 6, 1-21.

Predicting mutagenicity using Sarah Nexus

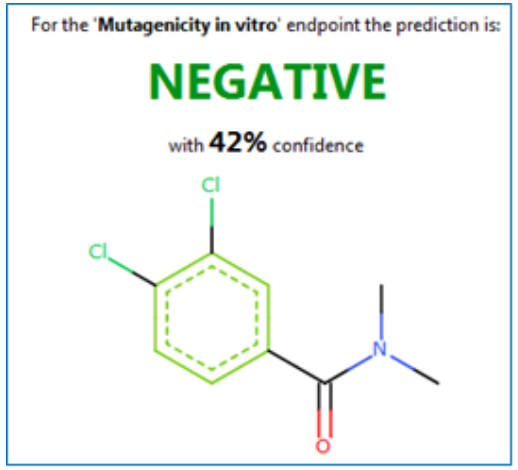
How Sarah works…

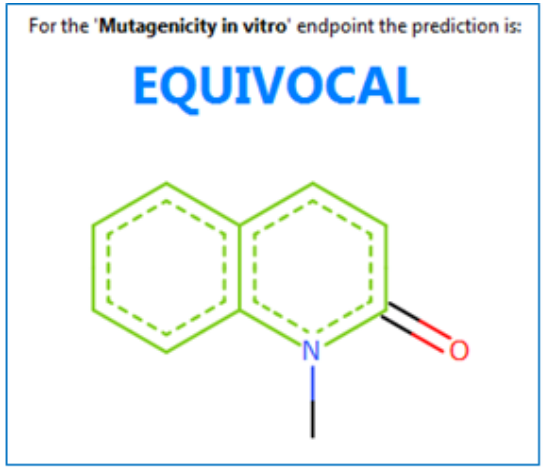
- A recursive decision tree identifies (de)activating fragments
- Each fragment forms a potential hypothesis
  - *If I see this fragment, I associate this with a positive or negative result in the Ames Test*
- Each hypothesis is assembled into a network
- Self Organising Hypothesis Network (SOHN)

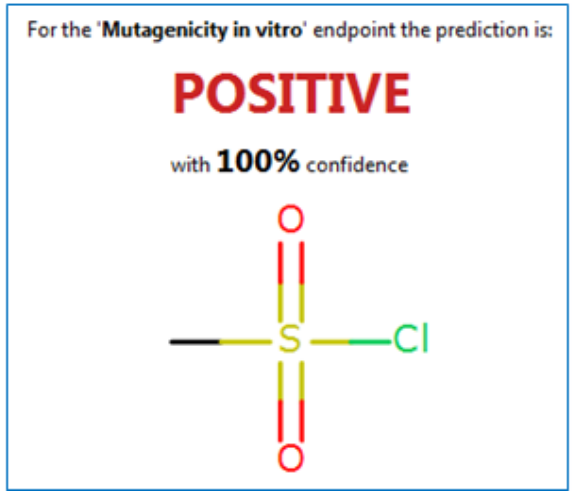


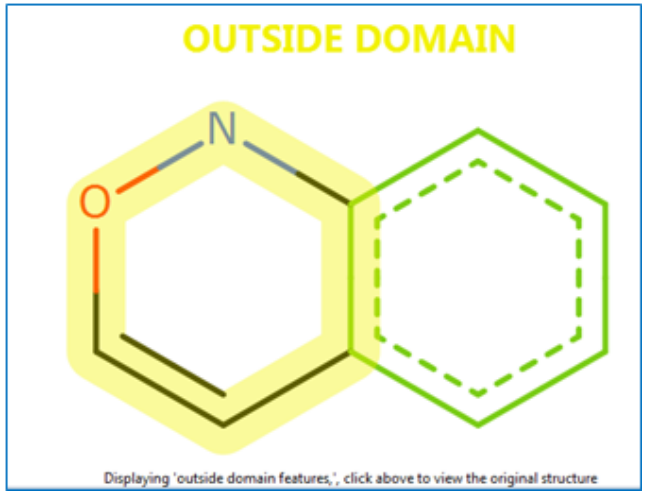
Hanser et al., *Journal of Cheminformatics*, 6, 1-21.

Predicting mutagenicity using Sarah Nexus





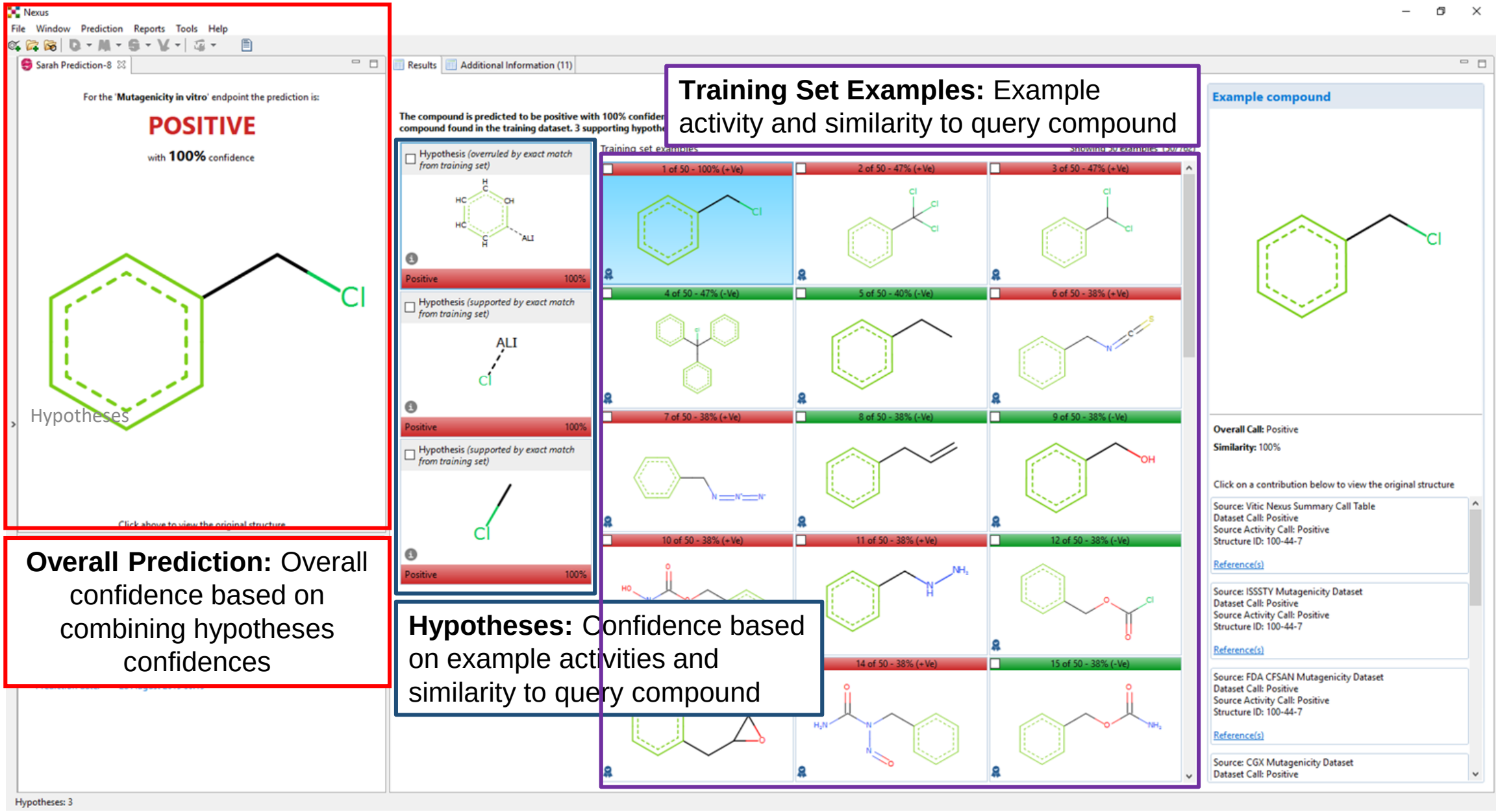




- Sarah has a defined applicability domain
- Predictions are transparent and therefore interpretable

Hanser et al., *Journal of Cheminformatics*, 6, 1-21.

Predicting mutagenicity using Sarah Nexus



Training Set Examples: Example activity and similarity to query compound

Overall Prediction: Overall confidence based on combining hypotheses confidences

Hypotheses: Confidence based on example activities and similarity to query compound

Comparison of expert rule-based and statistical-based systems

Table 2 Strengths and weaknesses of different *in silico* approaches

Method	*Strengths*	*Weaknesses*
Statistical-based	- Assesses all areas of the molecule, including whole molecular properties - Can often explain the structural basis for the prediction - Quick to generate a model - Fast to compute a prediction	- No direct mechanism provided, but may be elucidated as part of the QSAR explanation - Difficult to model complex *in vivo* outcomes
Expert rule-based	- Capable of precisely identifying reactive groups or pharmacophores that are linked to a mechanism - Includes reasons for activation and deactivation - Fast to compute a prediction	- Usually localized to the presence of a substructure and often does not take into account the whole molecule - Difficult to make supported negative predictions

Table from: Myatt, G. J., Beilke, L. D., Cross, K. P. (2017) **In Silico Tools and their Application.** In: Samuel Chackalamannil, David P Rotella and Simon E Ward, (eds.) Comprehensive Medicinal Chemistry III vol. 4, pp. 156–176. Oxford: Elsevier. http://dx.doi.org/10.1016/B978-0-12-409547-2.12379-0

Expert review and predicting toxicity

Computational Toxicology
Specialty Section

Expert review

- An expert review is prudent to increase the reliability of an *in silico* prediction
- *In silico* systems provide additional supporting information
  - *Expert rule-based methodologies*: prediction confidence, structural basis, mechanistic basis, examples, underlying study information, …
  - *Statistical-based methodologies*: structural descriptors, phys-chem descriptors, confidence metric, contributions, underlying data, study information, information on analogs, …
- An expert review involves an examination of this supporting data alongside any other relevant knowledge to
  - Provide rationale to support the *in silico* conclusion
  - To support assessments of inconclusive predictions (e.g., equivocal predictions, out-of-domain assessments)
  - Potentially refuting a prediction with sufficient evidence
- Expert review is complete when the user is satisfied that the prediction is correct

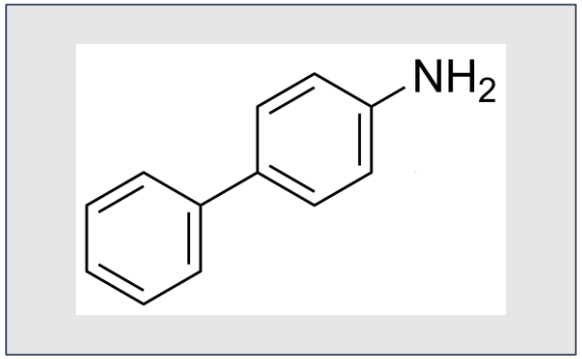
Expert review

- Expert review is recommended for certain regulations e.g. ICH M7 (DNA-reactive mutagenicity).
- When both systems predict the same outcome the probability of the predictions being correct is high.
- Expert review is prudent when the outcomes from the two complementary systems differ or there is some uncertainty.

Statistical-based	Expert rule-based	Percent mutagens	Expert review
Negative	Negative	< 10%	Likely to conclude negative
OOD	Negative or OOD	10%-20%	Expert review proportional to the likelihood of missing a mutagenic impurity
Negative	IND		
IND	Negative or IND	20%-30%	
Positive	IND or Negative		
OOD	IND		
OOD or Negative	Positive	30%-40%	
Positive or IND	Positive	> 50%	Likely to conclude positive

Table summarized from: Amberg, A., et al., 2019. **Principles and procedures for handling out-of-domain and indeterminate results as part of ICH M7 recommended (Q)SAR analyses.** Regulatory Toxicology and Pharmacology 102, 53–64. doi:10.1016/j.yrtph.2018.12.007

Predicting toxicity - Example 1

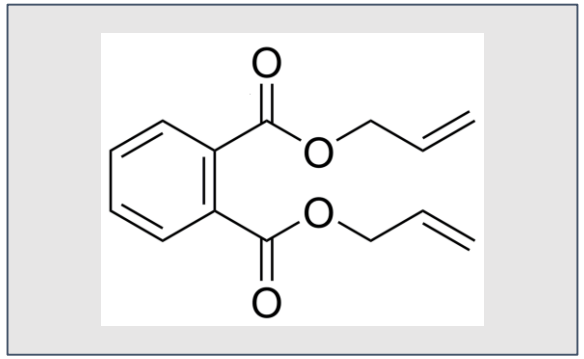


	Expert rule-based system	Statistical model
Lhasa software	**probable**	**positive (100% confidence)**
Leadscope, Inc. software	**positive**	**positive**

Supporting information

- Aromatic amines well-known to be mutagenic in the Ames test
- Mechanism thought to go through a reactive nitrenium ion which can bind to DNA
- Positive Ames data available for this exact compound in multiple datasets

Predicting toxicity - Example 2

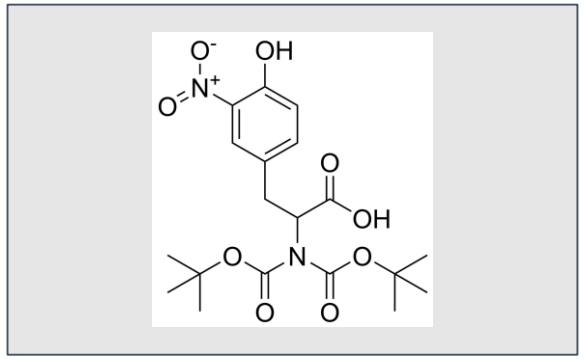


	Expert rule-based system	Statistical model
Lhasa software	**inactive**	**negative (100% confidence)**
Leadscope, Inc. software	**negative**	**negative**

Supporting information

- No alerting (reactive) features present on this compound
- Negative Ames data available for this exact compound in multiple datasets

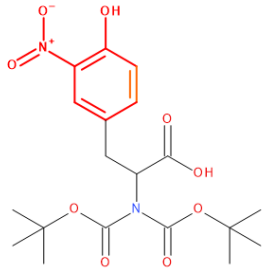
Predicting toxicity - Example 3



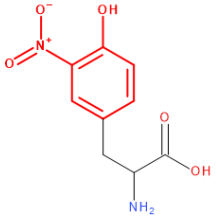
	Expert rule-based system	Statistical model
Lhasa software	**inactive**	**outside domain**
Leadscope, Inc. software	**indeterminate**	**negative**

Supporting information

- Chemical analog supports a negative assessment



Query chemical



Negative chemical

Key learnings

- Defined *in silico* toxicology and current use cases
- Defined main *in silico* methodologies
- Explained how expert rule-based and statistical-based methodologies are developed and used to make predictions
- Explained the role of expert review
- Provided examples of how predictions are made

Conclusions

- *In silico* toxicology is a fast and inexpensive approach to support toxicological assessments
- It is accepted as part of regulatory submissions (e.g., ICH M7)
- Using multiple methodologies alongside an expert review represents current best practice in the field

Reading list

General introduction to *in silico* methods and methodologies

Myatt, G. J., Beilke, L. D., Cross, K. P. (2017) **In Silico Tools and their Application.** In: Samuel Chackalamannil, David P Rotella and Simon E Ward, (eds.) Comprehensive Medicinal Chemistry III vol. 4, pp. 156–176. Oxford: Elsevier. http://doi.org/10.1016/B978-0-12-409547-2.12379-0

Williams R., et al. (2019) **Modelling Simple Toxicity Endpoints: Alerts, (Q)SARs and Beyond.** In: Hong H. (eds) Advances in Computational Toxicology. Challenges and Advances in Computational Chemistry and Physics, vol 30. Springer, Cham. https://doi.org/10.1007/978-3-030-16443-0_3

Hanser, T., et al., 2019. **Applicability Domain: Towards a More Formal Framework to Express the Applicability of a Model and the Confidence in Individual Predictions.** In: Hong H. (eds) Advances in Computational Toxicology. Challenges and Advances in Computational Chemistry and Physics, vol 30. Springer, Cham. https://doi.org/10.1007/978-3-030-16443-0_11

Regulatory (Q)SAR related document

ECHA, 2016. **Practical Guide: How to Use and Report (Q)SARs.** https://echa.europa.eu/documents/10162/13655/pg_report_qsars_en.pdf.

ECHA, 2017. **Read-across Assessment Framework (RAAF).** https://echa.europa.eu/documents/10162/13628/raaf_en.pdf.

EFSA, 2016. **Guidance on the establishment of the residue definition for dietary risk assessment: EFSA Panel on Plant Protection Products and their Residues (PPR).** EFSA J. 14, 1–12. http://onlinelibrary.wiley.com/doi/10.2903/j.efsa.2016.4549/epdf.

OECD, 2007. **Guidance Document on the Validation of (Quantitative) Structure-activity Relationships [(Q)SAR] Models,** OECD Environment Health and Safety Publications. Series on Testing and Assessment No. 69. http://www.oecd.org/env/guidance-document-on-the-validation-of-quantitative-structureactivity-relationship-q-sar-models-9789264085442-en.htm.

TSCA, 2016. **Toxic Substances Control Act (TSCA).** https://www.congress.gov/bill/114th-congress/senate-bill/697/all-info.

Expert review considerations

Powley, M.W., 2015. **(Q)SAR assessments of potentially mutagenic impurities: a regulatory perspective on the utility of expert knowledge and data submission.** Regul. Toxicol. Pharmacol. 71, 295-300. https://doi.org/10.1016/j.yrtph.2014.12.012

Barber, C., et al., 2015. **Establishing best practice in the application of expert review of mutagenicity under ICH M7.** Regulatory Toxicology and Pharmacology 73, 367–377. https://doi.org/10.1016/j.yrtph.2015.07.018

Amberg, A., et al., 2016. **Principles and procedures for implementation of ICH M7 recommended (Q)SAR analyses.** Regulatory Toxicology and Pharmacology 77, 13–24. https://doi.org/10.1016/j.yrtph.2016.02.004

Amberg, A., et al., 2019. **Principles and procedures for handling out-of-domain and indeterminate results as part of ICH M7 recommended (Q)SAR analyses.** Regulatory Toxicology and Pharmacology 102, 53–64. https://doi.org/10.1016/j.yrtph.2018.12.007

Protocols

Myatt, G.J., et al., 2018. ***In silico* toxicology protocols.** Regulatory Toxicology and Pharmacology 96, 1–17. https://doi.org/10.1016/j.yrtph.2018.04.014

Hasselgren, C.,et al., 2019. **Genetic toxicology *in silico* protocol.** Regulatory Toxicology and Pharmacology 107, 104403. https://doi.org/10.1016/j.yrtph.2019.104403

Hanser, T., et al., 2014. **Self organising hypothesis networks: a new approach for representing and structuring SAR knowledge.** Journal of Cheminformatics, 6, 1-21. https://doi.org/10.1186/1758-2946-6-21

SOT | Society of Toxicology

Creating a Safer and Healthier World by Advancing the Science and Increasing the Impact of Toxicology

Computational Toxicology Specialty Section

Thanks for listening

Please get in touch if you have any questions:

Donna Macmillan, Lhasa Limited: donna.macmillan@lhasalimited.org

Glenn Myatt, Leadscope, Inc: gmyatt@leadscope.com

Thank you SOT and CTSS!

SOT | Society of Toxicology

Creating a Safer and Healthier World by Advancing the Science and Increasing the Impact of Toxicology

Computational Toxicology Specialty Section

WEBINARS AVAILABLE ON CTSS WEBSITE!

Not an CTSS member yet? Consider joining today!

Joining is as easy as going to our website:
http://www.toxicology.org/groups/ss/ctss/
and following the links to "join CTSS"

Next CTSS Webinar (2 of 4):

save the date®

January 16, 2020

Title: US FDA Experience in the Regulatory Application of (Q)SAR Modeling

Speakers: Naomi Kruhlak, FDA CDER

Time: 11am-noon EST (1 hour)

CTSS Executive Committee

President
Lisa Beilke PhD DABT ERT
Toxicology Solutions

Vice-President
David Szabo PhD DABT
PPG

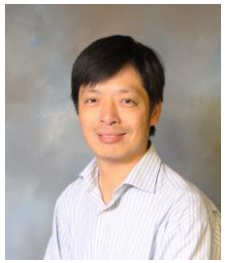
Vice-Pres Elect
Catrin Hasselgren PhD
Genentech

Past President
David Woolley PhD DABT ERT
ForthTox

Treasurer
Mark Gosink PhD
Pfizer

Secretary
Joel Cohen ScD DABT
Gradient

Councilor
Nigel Greene PhD
Astra Zeneca



Councilor
Minjun Chen PhD
FDA

Student Rep
Linlin Zhao PhD cand.
Rutgers University