Lhasa Limited

Release Notes

Lhasa Knowledge Suite - Nexus 2.5

Leaders in the development of expert chemoinformatic systems and trusted curators of proprietary data.

Statement of Intended Purpose

The computer programs in the Lhasa Knowledge Suite are intended to be used within a structured decision support system as part of the user's overall risk strategy.

Limited Warranty

Lhasa Limited makes no warranties, either expressed or implied, regarding the software described in this document or the online help, its merchantability, or its fitness for any particular purpose. In no event will Lhasa Limited be liable for any special, consequential, indirect or similar damages including any loss of profits or lost data arising out of the use of the software or data described in this document.

Trademarks

Acknowledgements

Lhasa Limited acknowledges the contributions to the following programs in the Lhasa Knowledge Suite:

For Derek Nexus:

- Members of the Collaborators Group
- Imperial Cancer Research Fund
- Judson Consulting Service
- Logic Programming Associates Limited
- City University (during the StAR project)
- Schering Agrochemicals Limited
- Harvard University
- BABEL developers

For Meteor Nexus:

- Meteor Steering Committee and User Group
- BABEL developers
- Patrik Rydberg, et al, University of Copenhagen (for SMARTCyp)

For Vitic Nexus:

- ITIC SAR Database project

For Sarah Nexus:

- Members who worked closely with us during the development and initial testing phases

Contact Details

Lhasa Limited	
Granary Wharf House	Reception ☎: +44 (0)113 394 6020
2 Canal Wharf	Applied Sciences ☎: +44 (0)113 394 6030
LEEDS	General ✉: info@lhasalimited.org
LS11 5PS	Applied Sciences ✉: hello@lhasalimited.org
United Kingdom	Website: www.lhasalimited.org

Lhasa Limited is a not-for-profit organisation.

Registered charity number 290866
(Registered in England and Wales)

Contents

What's New in Nexus 2.5 (Software)

Nexus 2.5 has the following new and improved program features. The changes are grouped by their patch number (2.5.X). Versions that do not contain any changes to functionality are not included in the Release Notes.

Nexus 2.5.2

EC3 Model Building Statistics

The EC3 model building statistics have been improved, giving clearer information on the dataset uploaded. The new statistics table displays:

- Compound entries in file
- Unique compounds in file
- Unique sensitisers (including any conflicted)
- Unique non-sensitisers
- Unique compounds with conflicting results
- Unique invalid compounds

Note on EC3 Models and Knowledge Base Update

Please note, as the Derek knowledge base has been updated, any EC3 models built in previous versions of Nexus 2.5 will not work. These can be removed through the UI.

M7C API Name on Reports

Previously the name of the API for an ICH M7 Classification prediction was not making its way to the Word reports. This has now been resolved.

M7C Long Compound Names in UI

In previous versions of Nexus, a very long compound name for any non-API compound could prevent the user from carrying out actions and performing their expert review. Now if a name is too long to display, the title shown is truncated with ellipsis in the middle of the word, preventing the UI from breaking. The full compound name is still preserved unchanged within the prediction, and can be found by hovering over the title of the **Edit results** page or within the **Name** input field.

Nexus 2.5.1

ICH M7 Class 4 Automatic Allocation when using Argument 5.1 (previously Argument 26)

In Nexus 2.5.0, the ICH M7 arguments have been re-numbered to be grouped and argument 26 was remapped to argument 5.1. Argument 26/5.1 is a negative call stating that:

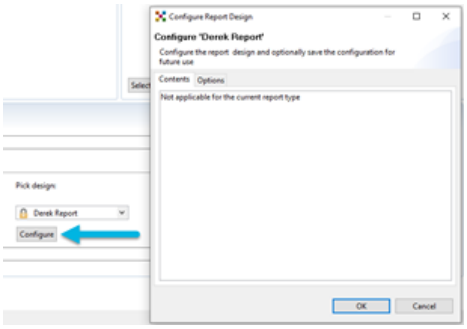
"Toxicophore identified by Derek Nexus is also present in the Ames negative API in the same chemical environment and there are no additional toxicophores present"

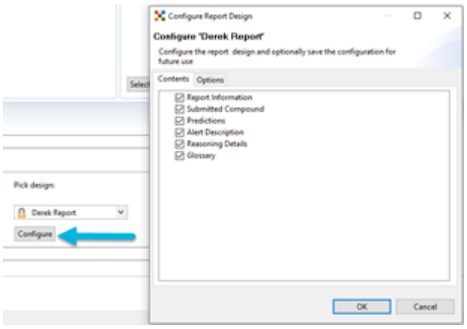
If selected during an ICH M7 Batch Classification expert review, this should result in an ICH M7 Class 4. However, it was erroneously being allocated as ICH M7 Class 5. This has been updated to provide the correct ICH M7 Class.

Known workaround for Nexus 2.5.0: Manually select ICH M7 Class 4 within the individual assessment.

Batch Reporting Customisation

Introduced in 2.5.0, the batch reporting configuration on batch setup pages was not populated (example below, on the left). This affected Derek, Sarah, and ICH M7 Batch predictions. The **Configure** window now functions correctly and allows for report configuration (example below, on the right).





Known workaround for Nexus 2.5.0: Create a custom report design in a single prediction report setup, save, and then select this custom design on the batch prediction setup screen.

Load of Legacy Predictions

Depending on the language setup of the computer, predictions generated and saved using previous versions of Nexus could no longer be loaded into Nexus 2.5.0. This predominantly affected locales using non-Latin characters such as Chinese.

Vitic Link Connection

Minor fix for Vitic Link relating to connections to Vitic 3 and Vitic 4 to avoid “Connection failed” message in some situations.

What's New in the Derek 2022 2.0 Knowledge Base (Knowledge)

New Alerts

Alert 956: To predict the skin sensitisation potential of uronium, halouronium or guanidinium salts.

Alert 957: To predict the skin sensitisation potential of activated phosphorous (V) compounds.

Scope Modification

Alert 477: Patterns were updated to prevent diketone compounds inadvertently alerting.

Knowledge Base Status

There are 940 alerts in the Derek 2022 2.0 Knowledge Base. The following table shows the number of enabled alerts for top level parent endpoints.

Endpoint	Number of Alerts
Carcinogenicity (ALL)	79
Genotoxicity (ALL) including Chromosome damage Mutagenicity	231 105 154
Irritation (ALL)	67
Miscellaneous endpoints (ALL)	116
Neurotoxicity (ALL)	10
Organ toxicity (ALL) including Hepatotoxicity	241 76
Reproductive toxicity (ALL)	61
Respiratory sensitisation (ALL)	13
Skin sensitisation (ALL)	133

What's New in Sarah Nexus 3.2.1 (Knowledge)

The following sections detail the changes to Sarah Nexus 3.2.1.

Sarah Model

A new Sarah Nexus model has been built called "Sarah Model – 2022.2", which uses Ames mutagenicity data from an expanded training set of 12,199 compounds sourced from data contained in the Vitic database and donated by Lhasa Limited members. The chemical structures of the training set compounds have been standardised and the biological data curated to reach an overall Ames result for each data point. This produces a self-organising hypothesis network (SOHN) model containing 418 hypotheses (289 unique).

Sarah training set data records

The training set to prepare Sarah Model – 2022.2 has been updated by the addition of Ames test data for new and existing chemicals within the model.

- Total data records = 50,692
- Total substances = 13,520
    - Number of positive substances = 5,814 (48%)
    - Number of negative substances = 6,385 (52%)
    - Number of additional substances = 1,321*

*Additional substances have not been included in the model but may be accessed in the Sarah **Additional information** tab. Although these are in the extended training set, they have been rejected by the model because they have neither a positive or negative overall call or for various other reasons during the standardisation process.

Sarah training set data sources

The following table shows the number of records contained in the Sarah training set from each data source.

Sources	Total number of records
ACID Halide Mutagenicity Dataset	39
Bursi Mutagenicity Dataset	4,326
Carcinogenic Potency Database (CPDB)	833
CGX Mutagenicity Dataset	709
Derek Example Compounds	363
EURL ECVAM Genotoxicity and Carcinogenicity Consolidated Database	956
Feng Mutagenicity Dataset	1,859
Hansen Mutagenicity Dataset	6,479
Helma Mutagenicity Dataset	683

Sources	Total number of records
ISSSTY Mutagenicity Dataset	7,051
Japan Chemical Industry Ecology-Toxicology and Information Center (JETOC) Mutagenicity Dataset	323
Marketed Pharmaceuticals Database	547
Member data	234
National Institute of Health Sciences Dataset	668
Vitic NTP Table	2,016
Vitic Summary call Table	14,899
US Food and Drug Administration - Center for Drug Evaluation and Research (FDA CDER)	585
US Food and Drug Administration - Center for Food Safety and Applied Nutrition (FDA CFSAN)	8,122
Total	**50,692**

Internal validation

5-Fold cross-validation is performed during the Sarah model building process and the results can be assessed within the **Manage prediction models** section of Nexus for “Sarah Model – 2022.2”.

External validation

External validation has been conducted against a selection of datasets containing proprietary Ames data. Performance statistics are provided for the predictions in the Library section of the Lhasa Limited website.

What's New in ICH M7 (Knowledge)

The following sections detail the changes to the ICH M7 function in Nexus 2.5.

ICH M7 Expert Review (Derek Nexus 6.2.1 and Sarah Nexus 3.2.1)

The following changes have been made to the automated expert review arguments displayed following an ICH M7 prediction.

Automated expert review arguments

Argument number 31.4 has been updated to align with changes made to Derek alert number 477 in Derek KB 2022 2.0.

Links between Derek Nexus and Sarah Nexus

Links between Derek alerts with Sarah hypotheses and training set examples have been updated using the latest knowledge base and model.

122 Sarah hypotheses have been linked to 61 Derek alerts.

- Sarah hypotheses may be linked to multiple Derek alerts
- Derek alerts may be linked to multiple Sarah hypotheses

6,182 Sarah training examples have been linked to 139 Derek alerts.

- Sarah training examples may be linked to multiple Derek alerts

ICH M7 Classification (Derek Nexus 6.2.1 and Sarah Nexus 3.2.1)

Th following changes have been made to the ICH M7 classification tool.

Ames Dataset

The Lhasa Ames dataset has been updated in line with the Sarah training set and the Mutagenicity Negative Predictions dataset.

Permissible Daily Exposure Comments

Comments are displayed in the ICH M7 classification table for queries which have had a permissible daily exposure (PDE) or acceptable intake (AI) or concentration limit published. These comments are displayed for any component of the query that matches a chemical in the published list of PDEs, AIs and concentration limits.

The following updates have been made to the information presented:

Chemical	CAS	Limit	Reference
Ethyl methanesulphonate	62-50-0	PDE = 1000 ug/day	Bercu et al, Regulatory Toxicology and Pharmacology, 2018, 94, 172-182
Benzene	71-43-2	Concentration limit = 2 ppm	ICH guideline Q3C (R8) on impurities: guidelines for residual solvents (May 2021)
Carbon tetrachloride	56-23-5	Concentration limit = 4 ppm	ICH guideline Q3C (R8) on impurities: guidelines for residual solvents (May 2021)

Chemical	CAS	Limit	Reference
1,1-Dichloroethene	75-35-4	Concentration limit = 8 ppm	ICH guideline Q3C (R8) on impurities: guidelines for residual solvents (May 2021)
1,1,1-Trichloroethane	71-55-6	Concentration limit = 1500 ppm	ICH guideline Q3C (R8) on impurities: guidelines for residual solvents (May 2021)
1,2-Dichloroethane	107-06-2	Concentration limit = 5 ppm	ICH guideline Q3C (R8) on impurities: guidelines for residual solvents (May 2021)

Derek Knowledge Base Changes in Previous Versions

The following changes were made in previous versions of the Derek 2022 Knowledge Base.

New and Modified Endpoints

New endpoint: Skin Sensitisation HPC

This endpoint contains alerts describing a skin sensitisation High Potency Category (HPC), consisting of a set of structural features within a mechanistic domain which are likely to be associated with extreme skin sensitisation potential in the Local Lymph Node Assay. These HPC rules were first described in the context of the Dermal Sensitisation Threshold (DST),[1] which is a Threshold of Toxicological Concern for skin sensitisation.[2,3] The presence or absence of alerts for skin sensitisation HPC provides information about which of the DSTs is the most appropriate to use.[4]

Modified endpoint: Irritation (of the Skin)

Endpoint renamed to Skin Irritation/Corrision.

New and Modified Rules

Rule 1607 implemented: If a chemical is known to give a corrosive response in a skin irritation study in the rabbit then it is considered certain that the chemical will cause skin irritation/corrosion in rabbits, probable in mammals other than the rabbit and impossible in bacteria.

Rule 1609 implemented: If a chemical is known to give an irritant response in a skin irritation study in rabbits then it is considered certain that the chemical will cause skin irritation/corrosion in rabbits, probable in mammals other than the rabbit and impossible in bacteria.

Rule 1611 implemented: If a chemical is known to be corrosive in a reconstructed human epidermis test method in humans then it is considered certain that the chemical will cause skin irritation/corrosion in humans, probable in mammals and impossible in bacteria.

Rule 1615 implemented: If a chemical is known reconstructed human epidermis test irritant in human then it is considered certain that the chemical will cause skin irritation/corrosion in human, probable in mammals and impossible in bacteria.

Rule 1616 implemented: If the chemical has a LogP value greater than 0.6 then in mammals the variable "Species dependent variable 43" is plausible.

Rule 1619 implemented: If a chemical is known to be corrosive in a membrane barrier test method in humans then it is considered certain that the chemical will cause skin irritation/corrosion in humans, probable in mammals and impossible in bacteria.

Rule 1632 implemented: If a chemical is known to be irritant in human patch test studies then it is considered certain that the chemical will cause skin irritation/corrosion in humans, probable in mammals and impossible in bacteria.

Rule 1645 implemented: If a chemical is known to give an irritant response in a skin irritation study in guinea pigs then it is considered certain that the chemical will cause skin irritation/corrosion in guinea pigs, probable in mammals other than the guinea pig and impossible in bacteria.

---

1. Roberts et al, Regul Toxicol Pharmacol (2015) 72, 683-93.
2. Safford et al, Regul Toxicol and Pharmacol (2011) 60, 218-224.
3. Safford et al, Regul Toxicol Pharmacol (2015), 72, 694-701.
4. Chilton et al, Regul Toxicol Pharmacol (2022), 133, 105200.

New and Improved Alerts

Carcinogenicity

Alert number	Status	Alert name / Chemical Class	Examples	Modifications
70	enabled	N-Nitro or N-nitroso compound	new: N-nitrosodiphenylamine	Alert description updated to address potency categorisation of nitrosamines. Alert scope restricted to exclude nitrosamines where the amine nitrogen is aromatic.
102	enabled	Aromatic nitroso compound	new: p-nitrosodiphenylamine, 1-nitrosonaphthalene	Alert scope updated to include aromatic N-nitroso compounds.
105	enabled	Aromatic nitro compound	new: 3-(5-nitro-2-furanyl)-imidazo[1,2-a]pyridine, metronidazole	Alert scope updated to include aromatic N-nitro compounds.
586	enabled	Aromatic amine or amide		Alert scope updated to align with the corresponding mutagenicity alert.
587	enabled	Aromatic amine or amide		Alert scope updated to align with the corresponding mutagenicity alert.
588	enabled	Aromatic amine or amide		Alert scope updated to align with the corresponding mutagenicity alert.

Chromosome damage *in vivo*

Alert number	Status	Alert name / Chemical Class	Examples	Modifications
7	enabled	N-Nitro or N-nitroso compound		Alert scope restricted to exclude nitrosamines where the amine nitrogen is aromatic and include acyclic diisopropylamine derivatives, based on public and proprietary Ames test data.

Alert number	Status	Alert name / Chemical Class	Examples	Modifications
329	enabled	Aromatic nitro compound		Alert scope updated to exclude esters, carbonates and carbamates in the ortho/para position, ortho-amino nitrobenzenes and compounds with a trifluoromethyl or sulphonyl substituent on the ring, based on proprietary in vitro mutagenicity data.
578	enabled	5-Fluoropyrimidine	new: emtricitabine	Alert updated to include the in vivo chromosome damage endpoint.
706	enabled	Alkyl ester of phosphoric or phosphonic acid	new: monocrotophos	Alert scope updated to include alkyl phosphonates and a wider range of phosphorothiolate esters.
952	enabled	Aromatic amine	new: 2-chloroaniline, 4-methyl-ortho-phenylenediamine, acetanilide; linked: 4-chloro-ortho-phenylenediamine, aniline	Alert implemented to predict the in vivo chromosome damage potential of aromatic amine compounds, based on public data.
953	enabled	4-Amino biphenyl, stilbene or derivative	new: 4,4''-diamino-para-terphenyl, 4-aminostilbene, 2,7-diaminofluorene; linked: benzidine, 2-acetamidofluorene, 4-aminobiphenyl	Alert implemented to predict the in vivo chromosome damage potential of 4-amino biphenyl or stilbene compounds and derivatives, based on public data.
954	enabled	Conjugated alkene	linked: 1,3-butadiene, isoprene	Alert implemented to predict the in vivo chromosome damage potential of conjugated alkene compounds, based on public data.

Chromosome damage *in vitro*

Alert number	Status	Alert name / Chemical Class	Examples	Modifications
7	enabled	N-Nitro or N-nitroso compound		Alert scope restricted to exclude nitrosamines where the amine nitrogen is aromatic and include acyclic diisopropylamine derivatives, based on public and proprietary Ames test data.
27	enabled	Alkylating agent		Alert scope restricted to exclude benzylchlorides with a halogen substituent on the benzene ring, based on proprietary Ames test data.
307	enabled	N-Methylol compound or precursor	new: N-(hydroxymethyl)methacrylamide, diazolidinyl urea, dimethyl [3-[(hydroxymethyl)amino]-3-oxopropyl]phosphonate	Alert scope updated to include tautomers of triazine precursors. Alert updated to include the in vitro chromosome damage endpoint.
311	enabled	alpha,beta-Unsaturated amide or thioamide		Alert updated to include the in vitro chromosome damage endpoint.

Alert number	Status	Alert name / Chemical Class	Examples	Modifications
329	enabled	Aromatic nitro compound		Alert scope updated to exclude esters, carbonates and carbamates in the ortho/para position, ortho-amino nitrobenzenes and compounds with a trifluoromethyl or sulphonyl substituent on the ring, based on proprietary in vitro mutagenicity data.
706	enabled	Alkyl ester of phosphoric or phosphonic acid	new: monocrotophos, phosphamidon	Alert scope updated to include alkyl phosphonates and a wider range of phosphorothiolate esters.
951	enabled	Oestradiol or analogue	linked: 17beta-oestradiol, ethynyloestradiol	Alert implemented to predict the in vitro chromosome damage potential of oestradiol compounds and analogues, based on public data.

Mutagenicity *in vitro*

Alert number	Status	Alert name / Chemical Class	Examples	Modifications
7	enabled	N-Nitro or N-nitroso compound		Alert scope restricted to exclude nitrosamines where the amine nitrogen is aromatic and include acyclic diisopropylamine derivatives, based on public and proprietary Ames test data.

Alert number	Status	Alert name / Chemical Class	Examples	Modifications
27	enabled	Alkylating agent		Alert scope restricted to exclude benzylchlorides with a halogen substituent on the benzene ring, based on proprietary in vitro mutagenicity data.
207	enabled	O/N-Substituted anthraquinone or precursor		Alert scope modified to include methoxy, amino, amido, nitro and diazonium substituents, based on public data.
302	enabled	alpha,beta-Unsaturated aldehyde or precursor		Alert scope restricted to exclude alpha,beta-unsaturated aldehydes which contain an oxygen substituent at the beta position and a carbon substituent at the alpha position, based on public data.
307	enabled	N-Methylol compound or precursor		Alert scope updated to include tautomers of triazine precursors. Alert updated to include the in vitro chromosome damage endpoint.
329	enabled	Aromatic nitro compound	unlinked: 2-nitrophenol, 4-nitrophenol	Alert scope updated to exclude esters, carbonates and carbamates in the ortho/para position, ortho-amino nitrobenzenes and compounds with a trifluoromethyl or sulphonyl substituent on the ring, based on proprietary in vitro mutagenicity data.
352	enabled	Aromatic amine or amide	unlinked: ropinirole, 3,4-dichloroaniline	Alert modified to include m-C=OOC substituted anilines, based on public and proprietary data.
353	enabled	Aromatic amine or amide		Alert scope modified to include certain 4-amino-1,2,3-triazoles, based on proprietary data.

Alert number	Status	Alert name / Chemical Class	Examples	Modifications
746	enabled	Arylboronic acid or derivative		Alert scope updated to exclude aryl boronic acids and derivatives with an electron donating group at the para-position, in the absence of an electron withdrawing group at the meta-position, and pyrazole-4-boronic acid derivatives bearing a ring N substitution that is not readily hydrolysable, based on proprietary data.
769	enabled	Anthracene		Alert scope restricted to exclude anthracenes with an aromatic substituent at C9, based on proprietary data.
946	enabled	beta- or gamma-Carboline	new: 4-amino-3-carboline, harmine hydrochloride	Alert implemented to predict the in vitro mutagenicity potential of beta- or gamma-carboline compounds, based on public data.

Mutagenicity *in vivo*

Alert number	Status	Alert name / Chemical Class	Examples	Modifications
7	enabled	N-Nitro or N-nitroso compound		Alert scope restricted to exclude nitrosamines where the amine nitrogen is aromatic and include acyclic diisopropylamine derivatives, based on proprietary Ames test data.
329	enabled	Aromatic nitro compound		Alert scope updated to exclude esters, carbonates and carbamates in the ortho/para position, ortho-amino nitrobenzenes and compounds with a trifluoromethyl or sulphonyl substituent on the ring, based on proprietary in vitro mutagenicity data.
352	enabled	Aromatic amine or amide		Alert modified to include m-C=OOC substituted anilines, based on public and proprietary in vitro mutagenicity data.

Irritation (of the eye)

Alert number	Status	Alert name / Chemical Class	Examples	Modifications
40	enabled	Nitrogen or sulphur mustard		Alert updated to no longer predict for the irritation (of the skin) endpoint.
43	enabled	alpha-Halo aliphatic ester		Alert updated to no longer predict for the irritation (of the skin) endpoint.
44	enabled	Acid halide		Alert updated to no longer predict for the irritation (of the skin) endpoint.
85	enabled	Hydrazine or monoacyl- or monosulphonyl-hydrazine		Alert updated to no longer predict for the irritation (of the skin) endpoint.
210	enabled	Acid anhydride		Alert updated to no longer predict for the irritation (of the skin) endpoint and to include the irritation (of the respiratory tract) endpoint.
211	enabled	Isocyanate		Alert updated to no longer predict for the irritation (of the skin) endpoint.
212	enabled	alpha,beta-Unsaturated ester		Alert updated to no longer predict for the irritation (of the skin) endpoint and to include the irritation (of the respiratory tract) endpoint.
213	enabled	Aluminium alkyl		Alert updated to no longer predict for the irritation (of the skin) endpoint and to include the irritation (of the respiratory tract) endpoint.
214	enabled	Boron alkyl		Alert updated to no longer predict for the irritation (of the skin) endpoint.
215	enabled	Boron hydride		Alert updated to no longer predict for the irritation (of the skin) endpoint.
216	enabled	Boron halide		Alert updated to no longer predict for the irritation (of the skin) endpoint.

Alert number	Status	Alert name / Chemical Class	Examples	Modifications
217	enabled	Silicon halide		Alert updated to no longer predict for the irritation (of the skin) endpoint.
218	enabled	Quaternary ammonium salt		Alert updated to no longer predict for the irritation (of the skin) endpoint.
219	enabled	Peracid		Alert updated to no longer predict for the irritation (of the skin) endpoint.
220	enabled	alpha,beta-Unsaturated aldehyde		Alert updated to no longer predict for the irritation (of the skin) endpoint.
221	enabled	Dialdehyde		Alert updated to no longer predict for the irritation (of the skin) endpoint and to include the irritation (of the respiratory tract) endpoint.
222	enabled	Epoxide		Alert updated to no longer predict for the irritation (of the skin) endpoint.
223	enabled	Dialkyl sulphate		Alert updated to no longer predict for the irritation (of the skin) endpoint.
224	enabled	Chromate or dichromate		Alert updated to no longer predict for the irritation (of the skin) endpoint and to include the irritation (of the respiratory tract) endpoint.
225	enabled	Hydrogen halide		Alert updated to no longer predict for the irritation (of the skin) endpoint.
227	enabled	Allyl halide		Alert updated to no longer predict for the irritation (of the skin) endpoint.
228	enabled	Benzyl halide		Alert updated to no longer predict for the irritation (of the skin) endpoint.
229	enabled	Alkyl hydroperoxide		Alert updated to no longer predict for the irritation (of the skin) endpoint.

Alert number	Status	Alert name / Chemical Class	Examples	Modifications
230	enabled	N-Haloimide		Alert updated to no longer predict for the irritation (of the skin) endpoint.
231	enabled	N-Halosulphonamide		Alert updated to no longer predict for the irritation (of the skin) endpoint.

Irritation (of the respiratory tract)

Alert number	Status	Alert name / Chemical Class	Examples	Modifications
40	enabled	Nitrogen or sulphur mustard		Alert updated to no longer predict for the irritation (of the skin) endpoint.
43	enabled	alpha-Halo aliphatic ester		Alert updated to no longer predict for the irritation (of the skin) endpoint.
44	enabled	Acid halide		Alert updated to no longer predict for the irritation (of the skin) endpoint.
85	enabled	Hydrazine or monoacyl- or monosulphonyl-hydrazine		Alert updated to no longer predict for the irritation (of the skin) endpoint.
210	enabled	Acid anhydride		Alert updated to no longer predict for the irritation (of the skin) endpoint and to include the irritation (of the respiratory tract) endpoint.
211	enabled	Isocyanate		Alert updated to no longer predict for the irritation (of the skin) endpoint.
212	enabled	alpha,beta-Unsaturated ester		Alert updated to no longer predict for the irritation (of the skin) endpoint and to include the irritation (of the respiratory tract) endpoint.
213	enabled	Aluminium alkyl		Alert updated to no longer predict for the irritation (of the skin) endpoint and to include the irritation (of the respiratory tract) endpoint.

Alert number	Status	Alert name / Chemical Class	Examples	Modifications
214	enabled	Boron alkyl		Alert updated to no longer predict for the irritation (of the skin) endpoint.
215	enabled	Boron hydride		Alert updated to no longer predict for the irritation (of the skin) endpoint.
216	enabled	Boron halide		Alert updated to no longer predict for the irritation (of the skin) endpoint.
217	enabled	Silicon halide		Alert updated to no longer predict for the irritation (of the skin) endpoint.
220	enabled	alpha,beta-Unsaturated aldehyde		Alert updated to no longer predict for the irritation (of the skin) endpoint.
221	enabled	Dialdehyde		Alert updated to no longer predict for the irritation (of the skin) endpoint and to include the irritation (of the respiratory tract) endpoint.
224	enabled	Chromate or dichromate		Alert updated to no longer predict for the irritation (of the skin) endpoint and to include the irritation (of the respiratory tract) endpoint.
225	enabled	Hydrogen halide		Alert updated to no longer predict for the irritation (of the skin) endpoint.
227	enabled	Allyl halide		Alert updated to no longer predict for the irritation (of the skin) endpoint.
228	enabled	Benzyl halide		Alert updated to no longer predict for the irritation (of the skin) endpoint.
229	enabled	Alkyl hydroperoxide		Alert updated to no longer predict for the irritation (of the skin) endpoint.
230	enabled	N-Haloimide		Alert updated to no longer predict for the irritation (of the skin) endpoint.

Alert number	Status	Alert name / Chemical Class	Examples	Modifications
231	enabled	N-Halosulphonamide		Alert updated to no longer predict for the irritation (of the skin) endpoint.

Skin irritation/corrosion

Alert number	Status	Alert name / Chemical Class	Examples	Modifications
913	enabled	Primary or secondary alkyl alcohol	new: (+/-) trans-3,3-dimethyl-5-(2,2,3-trimethyl-cyclopent-3-en-1-yl)pent-4-en-2-ol, cyclohexanol; linked: 2-ethylhexanol	Alert implemented to predict the skin irritation/corrosion potential of primary or secondary alkyl alcohol compounds, based on public data.
914	enabled	Alkyl aldehyde or precursor	new: 1,1-dimethoxyoctane, 3-methylbutanal, glutaraldehyde, heptanal; linked: 3-(4-isopropyl-phenyl)-2-methylpropionaldehyde, glyoxal	Alert implemented to predict the skin irritation/corrosion potential of alkyl aldehyde compounds or precursors, based on public data.
915	enabled	Phenol	new: 2-methoxyphenol, 4-chloro-o-cresol; linked: thymol	Alert implemented to predict the skin irritation/corrosion potential of phenol compounds, based on public data.
916	enabled	Styrene or derivative	new: alpha-methyl styrene	Alert implemented to predict the skin irritation/corrosion potential of styrene compounds or derivatives, based on public data.

Alert number	Status	Alert name / Chemical Class	Examples	Modifications
917	enabled	Alkyltin compound	new: butyltin trichloride, methyl (Z,Z)-8,8-dibutyl-3,6,10-trioxo-2,7,9-trioxa-8-stannatrideca-4,11-dien-13-oate, tributyltin chloride	Alert implemented to predict the skin irritation/corrosion potential of hydrogen halide compounds, based on public data.
918	enabled	Alkyl amine	new: 1,1,3,3-tetramethylbutylamine, 1-methylpyrrolidine, butylamine, dimethylamine, heptylamine, isopropylamine, morpholine, triethylamine	Alert implemented to predict the skin irritation/corrosion potential of alkyl amine compounds, based on public data.
919	enabled	Trihalo or tetrahaloalkene	linked: tetrachloroethylene, trichloroethylene	Alert implemented to predict the skin irritation/corrosion potential of trihalo or tetrahaloalkenyl compounds, based on public data.
920	enabled	Chromate or dichromate	linked: chromium trioxide, potassium dichromate	Alert implemented to predict the skin irritation/corrosion potential of chromate or dichromate compounds, based on public data.
921	enabled	Short chain alkyl carboxylic acid	new: mercaptoacetic acid, methoxyacetic acid, octanoic acid, pivalic acid	Alert implemented to predict the skin irritation/corrosion potential of short chain alkyl carboxylic acid compounds, based on public data.

Alert number	Status	Alert name / Chemical Class	Examples	Modifications
922	enabled	Alkyl halide	new: 1-bromohexane, 3-bromopropene, ethyl 2-chloroacetoacetate, iodomethane; linked: dichloromethane	Alert implemented to predict the skin irritation/corrosion potential of alkyl halide compounds, based on public data.
923	enabled	Acid anhydride	new: acetic anhydride, isobutyric anhydride, maleic anhydride, succinic anhydride	Alert implemented to predict the skin irritation/corrosion potential of acid anhydride compounds, based on public data.
924	enabled	Hypohalous acid or precursor	linked: sodium chlorite	Alert implemented to predict the skin irritation/corrosion potential of hypohalous acid compounds or precursors, based on public data.
925	enabled	Quaternary ammonium salt	new: cetylpyridinium chloride, stearyl trimethyl ammonium chloride, tricyclo[3.3.1.13,7]decan-1-aminium, N,N,N-trimethyl, hydroxide (1:1)	Alert implemented to predict the skin irritation/corrosion potential of quaternary ammonium salt compounds, based on public data.
926	enabled	Hydrogen halide	new: hydrogen chloride, hydrogen fluoride	Alert implemented to predict the skin irritation/corrosion potential of alkyltin compounds, based on public data.

Alert number	Status	Alert name / Chemical Class	Examples	Modifications
927	enabled	Phosphoric or phosphonic acid or derivative	new: orthophosphoric acid, phosphonic acid; linked: dibutyl phosphate	Alert implemented to predict the skin irritation/corrosion potential of phosphoric or phosphonic acid compounds or derivatives, based on public data.
928	enabled	Cyclohexene or derivative	new: 5-ethylidene-2-norbornene, cyclohexene, dicyclopentadiene, vinyl norbornene	Alert implemented to predict the skin irritation/corrosion potential of cyclohexenyl compounds or derivatives, based on public data.
929	enabled	Borohydride	new: sodium borohydride	Alert implemented to predict the skin irritation/corrosion potential of borohydride compounds, based on public data.
930	enabled	Nitrobenzene	new: 2,5-dinitrotoluene, 4-nitrotoluene-2-sulfonic acid	Alert implemented to predict the skin irritation/corrosion potential of nitrobenzene compounds, based on public data.
931	enabled	Monoalkyl or alkylaryl sulphate or sulphonate	new: dodecylbenzenesulphonic acid, sodium 1,4-bis(1,3-dimethylbutyl) sulphonatosuccinate, sodium dodecyl sulphate, sodium octyl sulphate	Alert implemented to predict the skin irritation/corrosion potential of monoalkyl or alkylaryl sulphate or sulphonate compounds, based on public data.

Alert number	Status	Alert name / Chemical Class	Examples	Modifications
932	enabled	Isocyanate or isothiocyanate	new: 2,4,6-triiospropyl-m-phenylene diisocyanate, 2,4,6-triiospropyl-m-phenylene diisocyanate, 3-(trimethoxysilyl)propyl isocyanate, alpha,alpha,alpha-trifluoro-3-tolyl isocyanate, tert-butyl isocyanate	Alert implemented to predict the skin irritation/corrosion potential of isocyanate or isothiocyanate compounds, based on public data.
933	enabled	Acid halide	new: 3-phenylpropionyl chloride, phosphorus tribromide; linked: dimethylsulphamoyl chloride, methylphosphonous dichloride, octanoyl chloride, pivaloyl chloride	Alert implemented to predict the skin irritation/corrosion potential of acid halide compounds, based on public data.
934	enabled	Terpenoid	new: 3,7-dimethylnona-1,6-dien-3-ol, 3,7-dimethylocta-1,3,6-triene, geranyl linalool; linked: linalool	Alert implemented to predict the skin irritation/corrosion potential of terpenoid compounds, based on public data.
935	enabled	Alkyl sulphate or sulphonate	new: 1,3,2-dioxathiolane 2,2-dioxide, methyl toluene-4-sulphonate; linked: 1,3-propane sultone, dimethyl sulphate	Alert implemented to predict the skin irritation/corrosion potential of alkyl sulphate or sulphonate compounds, based on public data.

Alert number	Status	Alert name / Chemical Class	Examples	Modifications
936	enabled	alpha,beta-Unsaturated aldehyde, ester or ketone	new: (E)-1-(2,6,6-trimethyl-1,3-cyclohexadien-1-yl)-2-buten-1-one, (E)-3-formylbut-2-enyl acetate, 2,4-decadienal, 2,5,6-trimethylcyclohex-2-en-1-one, 6,10-dimethylundeca-3,5,9-trien-2-one, alpha-hexylcinnamic aldehyde, tert-butyl methacrylate, tetrahydrofurfuryl acrylate; linked: 2-methyl-2-propenal (methacrolein), crotonaldehyde	Alert implemented to predict the skin irritation/corrosion potential of alpha,beta-unsaturated aldehyde, ester or ketone compounds, based on public data.
938	enabled	Hydroperoxide or peracid	new: tert-pentyl hydroperoxide; linked: peracetic acid	Alert implemented to predict the skin irritation/corrosion potential of hydroperoxide or peracid compounds, based on public data.
939	enabled	Borane or alkyl borane	new: borane-tetrahydrofuran complex	Alert implemented to predict the skin irritation/corrosion potential of borane or alkyl borane compounds, based on public data.
940	enabled	Pyridine or analogue	new: 2-methylpyridine, 3,5-dimethylpyridine, 5-ethyl-2-methylpyridine, isonicotinaldehyde, methyl 3-pyridyl ketone, methyl nicotinate, nicotine	Alert implemented to predict the skin irritation/corrosion potential of pyridine compounds or analogues, based on public data.
941	enabled	Epoxide	new: 1,4-bis[(2,3-epoxypropoxy)methyl]cyclohexane, 2,2-dimethyloxirane, octyloxirane	Alert implemented to predict the skin irritation/corrosion potential of epoxide compounds, based on public data.

Alert number	Status	Alert name / Chemical Class	Examples	Modifications
942	enabled	N-Halo ammonia, amide, imide or sulphonamide	new: monochloramine, sodium N-chlorobenzenesulphonamide, symclosene; linked: chloramine-T	Alert implemented to predict the skin irritation/corrosion potential of N-halo ammonia, amide, imide or sulphonamide compounds, based on public data.
943	enabled	Boron halide	new: boron trifluoride dihydrate, sodium tetrafluoroborate	Alert implemented to predict the skin irritation/corrosion potential of boron halide compounds, based on public data.
944	enabled	Silicon halide	new: chlorotrimethylsilane, dichlorosilane, silicon tetrachloride	Alert implemented to predict the skin irritation/corrosion potential of silicon halide compounds, based on public data.
945	enabled	Hydrazine or derivative	linked: hydrazine, methylhydrazine, phenylhydrazine	Alert implemented to predict the skin irritation/corrosion potential of hydrazine compounds or derivatives, based on public data.

Skin sensitisation

Alert number	Status	Alert name / Chemical Class	Examples	Modifications
425	enabled	Enol ether		Alert scope restricted to exclude vinyl esters, now covered elsewhere in the knowledge base.

Alert number	Status	Alert name / Chemical Class	Examples	Modifications
435	enabled	Diamine	new: HC orange no 2; linked: chlorpromazine	Alert scope updated to include tertiary amines and aromatic amines, based on public data.
442	enabled	Dithiocarbamate derivative	new: O,S-dibutyl dithioimidodicarbonate, dipentamethylenethiuram tetrasulphide	Alert scope updated to include dithiocarbamate esters and thiuram tri- and tetrasulphides, based on public data.
452	enabled	Allyl or alkyl peroxide	new: 1,1,3,3-tetramethylbutyl hydroperoxide, ascaridole	Alert scope updated to include alkyl peroxides, based on public data.
480	enabled	alpha,beta-Unsaturated ketone or precursor	unlinked: (5R)-2,3-dimethyl-5-isopropenyl-2-cyclohexene-1-one	
481	enabled	alpha,beta-Unsaturated ester or precursor	new: ethyl beta-safranate	Alert scope updated to include extended conjugation, based on public data.
837	enabled	Amino- or hydroxy-substituted aniline or aminopyridine	new: 2,6-diamino pyridine, 2,6-dimethoxy-3,5-pyridinediamine dihydrochloride, 6-methoxy-2-methylamino-3-aminopyridine dihydrochloride	Alert scope updated to include pyridines, based on proprietary data.
947	enabled	Benzylic, allylic or propargylic ester, carbamate, carbonate or phenolate	new: 2-propenyl octanoate, allyl phenoxyacetate, iodopropynyl butylcarbamate	Alert implemented to predict the skin sensitisation potential of benzylic, allylic or propargylic ester, carbamate, carbonate or phenolate compounds, based on public data.

Alert number	Status	Alert name / Chemical Class	Examples	Modifications
948	enabled	4- or 5-Aminopyrazole	new: 2-((3-aminopyrazolo (1,5-a)pyridin-2-yl)oxy)ethanol hydrochloride, 2-methyl-pyrazolo(5,1-b)quinazolin-9-one, 4,5-diamino-1-hexyl-1H-pyrazole dihydrochloride	Alert implemented to predict the skin sensitisation potential of 4- or 5-aminopyrazole compounds, based on public data.
949	enabled	Vinyl ester, carbonate or carbamate	new: 3-propylidenephthalide, vinyl 2-ethylhexanoate	Alert implemented to predict the skin sensitisation potential of vinyl ester, carbonate or carbamate compounds, based on public and proprietary data.
950	enabled	Aromatic primary hydroxymethyl compound	new: (pyridin-3-yl)methanol hydrofluoride, 4-methoxybenzyl alcohol; linked: furfuryl alcohol	Alert implemented to predict the skin sensitisation potential of aromatic primary hydroxymethyl compounds, based on public data.

Skin sensitisation HPC

Alert number	Status	Alert name / Chemical Class	Examples	Modifications
HPC01_1	enabled	Class 1: Protein derivatising agents		Alert implemented to describe the skin sensitisation High Potency Category (HPC) Class 1: Protein derivatising agents, as defined by Roberts et al 2015.
HPC02_1	enabled	Class 2: Michael acceptor		Alert implemented to describe the skin sensitisation High Potency Category (HPC) Class 2: Michael acceptor, as defined by Roberts et al 2015.
HPC02_2	enabled	Class 2: Michael acceptor		Alert implemented to describe the skin sensitisation High Potency Category (HPC) Class 2: Michael acceptor, as defined by Roberts et al 2015.

Alert number	Status	Alert name / Chemical Class	Examples	Modifications
HPC03_1	enabled	Class 3: Pro/pre-Michael acceptor		Alert implemented to describe the skin sensitisation High Potency Category (HPC) Class 3: Pro/pre-Michael acceptor, as defined by Roberts et al 2015.
HPC04_1	enabled	Class 4: Schiff base electrophiles		Alert implemented to describe the skin sensitisation High Potency Category (HPC) Class 4: Schiff base electrophiles, as defined by Roberts et al 2015.
HPC04_2	enabled	Class 4: Schiff base electrophiles		Alert implemented to describe the skin sensitisation High Potency Category (HPC) Class 4: Schiff base electrophiles, as defined by Roberts et al 2015.
HPC04_3	enabled	Class 4: Schiff base electrophiles		Alert implemented to describe the skin sensitisation High Potency Category (HPC) Class 4: Schiff base electrophiles, as defined by Roberts et al 2015.
HPC04_4	enabled	Class 4: Schiff base electrophiles		Alert implemented to describe the skin sensitisation High Potency Category (HPC) Class 4: Schiff base electrophiles, as defined by Roberts et al 2015.
HPC04_5	enabled	Class 4: Schiff base electrophiles		Alert implemented to describe the skin sensitisation High Potency Category (HPC) Class 4: Schiff base electrophiles, as defined by Roberts et al 2015.
HPC04_6	enabled	Class 4: Schiff base electrophiles		Alert implemented to describe the skin sensitisation High Potency Category (HPC) Class 4: Schiff base electrophiles, as defined by Roberts et al 2015.
HPC04_7	enabled	Class 4: Schiff base electrophiles		Alert implemented to describe the skin sensitisation High Potency Category (HPC) Class 4: Schiff base electrophiles, as defined by Roberts et al 2015.
HPC04_8	enabled	Class 4: Schiff base electrophiles		Alert implemented to describe the skin sensitisation High Potency Category (HPC) Class 4: Schiff base electrophiles, as defined by Roberts et al 2015.

Alert number	Status	Alert name / Chemical Class	Examples	Modifications
HPC05_1	enabled	Class 5: Acyl transfer agents		Alert implemented to describe the skin sensitisation High Potency Category (HPC) Class 5: Acyl transfer agents, as defined by Roberts et al 2015.
HPC05_2	enabled	Class 5: Acyl transfer agents		Alert implemented to describe the skin sensitisation High Potency Category (HPC) Class 5: Acyl transfer agents, as defined by Roberts et al 2015.
HPC06_1	enabled	Class 6: SN2 electrophiles		Alert implemented to describe the skin sensitisation High Potency Category (HPC) Class 6: SN2 electrophiles, as defined by Roberts et al 2015.
HPC06_2	enabled	Class 6: SN2 electrophiles		Alert implemented to describe the skin sensitisation High Potency Category (HPC) Class 6: SN2 electrophiles, as defined by Roberts et al 2015.
HPC06_3	enabled	Class 6: SN2 electrophiles		Alert implemented to describe the skin sensitisation High Potency Category (HPC) Class 6: SN2 electrophiles, as defined by Roberts et al 2015.
HPC07_1	enabled	Class 7: SNAr electrophiles		Alert implemented to describe the skin sensitisation High Potency Category (HPC) Class 7: SNAr electrophiles, as defined by Roberts et al 2015.
HPC08_1	enabled	Class 8: Organic peroxides		Alert implemented to describe the skin sensitisation High Potency Category (HPC) Class 8: Organic peroxides, as defined by Roberts et al 2015.

Sarah Nexus Changes in Previous Versions

The following changes were made in previous Nexus 2.5 releases.

Sarah Model

A new Sarah Nexus model has been built called “Sarah model – 2022.1”, which uses Ames mutagenicity data from an expanded training set of 12,195 compounds sourced from data contained in the Vitic database and donated by Lhasa Limited members. The chemical structures of the training set compounds have been standardised and the biological data curated to reach an overall Ames result for each data point. This produces a self-organising hypothesis network (SOHN) model containing 398 hypotheses (305 unique).

Sarah Training Set Data Records

The training set to prepare Sarah model – 2022.1 has been updated by the addition of Ames test data for several new and existing chemicals within the model.

- New data records = 4,607
- Total data records = 50,682
- New substances = 382
- Total substances = 13,519
    - Number of positive substances = 5,814 (48%)
    - Number of negative substances = 6,381 (52%)
    - Number of additional substances = 1,324*

*Additional substances have not been included in the model but may be accessed in the **Additional information** tab. Although these are in the extended training set, they have been rejected by the model because they have neither a positive or negative overall call or for various other reasons during the standardisation process.

Sarah Training Set Data Sources

The following table shows the number of records contained in the Sarah training set from each data source.

Sources	Total number of records
ACID Halide Mutagenicity Dataset	39
Bursi Mutagenicity Dataset	4,326
Carcinogenic Potency Database (CPDB)	833
CGX Mutagenicity Dataset	709
Derek Example Compounds	373
EURL ECVAM Genotoxicity and Carcinogenicity Consolidated Database	956
Feng Mutagenicity Dataset	1,859

Sources	Total number of records
Hansen Mutagenicity Dataset	6,479
Helma Mutagenicity Dataset	683
ISSSTY Mutagenicity Dataset	7,042
Japan Chemical Industry Ecology-Toxicology and Information Center (JETOC) Mutagenicity Dataset	323
Marketed Pharmaceuticals Database	547
Member data	232
National Institute of Health Sciences Dataset	668
Vitic NTP Table	2,016
Vitic Summary call Table	14,891
US Food and Drug Administration - Center for Drug Evaluation and Research (FDA CDER)	585
US Food and Drug Administration - Center for Food Safety and Applied Nutrition (FDA CFSAN)	8,121
Total	50,682

Internal Validation

5-Fold cross-validation is performed during the Sarah model building process and the results can be assessed within the **Manage prediction models** section of Nexus for “Sarah model – 2022.1”.

External Validation

6 datasets have been used for external validation:

- DGM/NIHS Dataset
- Proprietary Dataset 1
- Proprietary Dataset 2
- Proprietary Dataset 3
- Proprietary Dataset 4
- Vitic Intermediates Dataset

DGM/NIHS Dataset

A collection of Ames test data for 12,140 chemicals compiled by The Division of Genetics and Mutagenesis, National Institute of Health Sciences (DMG/NIHS) used in the Ames/QSAR International Challenge Project discussed in the following reference: Honma M et al. Improvement of quantitative structure-activity relationship (QSAR) tools for predicting Ames mutagenicity: outcomes of the Ames/QSAR International Challenge Project. Mutagenesis, 2019, 34, 3-16 available at

http://dx.doi.org/10.1093/mutage/gey031. Overall, 1,757 (14.4%) compounds have been assigned positive and 10383 (85.6%) have been assigned negative.

Proprietary Dataset 1

A proprietary collection of Ames test data for 454 pharmaceutical-related chemicals donated by a Lhasa Limited member. Overall, 54 (11.9%) compounds have been assigned positive and 400 (81.1%) have been assigned negative.

Proprietary Dataset 2

A proprietary collection of Ames test data for 507 pharmaceutical-related chemicals donated by a Lhasa Limited member. Overall, 95 (18.7%) compounds have been assigned positive and 412 (81.3%) have been assigned negative.

Proprietary Dataset 3

A proprietary collection of Ames test data for 2,865 pharmaceutical-related chemicals donated by a Lhasa Limited member. Overall, 171 (6.0%) compounds have been assigned positive and 2,694 (94.0%) have been assigned negative.

Proprietary Dataset 4

A proprietary collection of Ames test data for 1,278 pharmaceutical-related chemicals donated by a Lhasa Limited member. Overall, 259 (20.3%) compounds have been assigned positive and 1,019 (79.7%) have been assigned negative.

Vitic Intermediates Dataset

A collection of Ames test data for 1,748 common intermediates shared within the Vitic Intermediates projects. Information about Vitic Intermediates is available at https://www.lhasalimited.org/Initiatives/vitic-intermediates.htm. Overall, 584 (33.4%) compounds have been assigned positive and 1,164 (66.6%) have been assigned negative.

Dataset Performance for External Validation

The following table shows a summary of the data in the datasets used for external validations, as well as performance statistics for Sarah model – 2022.1 against each dataset.

Dataset					Performance Statistics											
Name	Size	Positive	Negative	Bias	BAC	SEN	SPEC	PPV	NPV	COV	TP	TN	FP	FN	EQ	OD
DMG/NIHS Dataset	12,140	1,757	10,383	-0.86	0.7587	0.7167	0.8006	0.3916	0.9404	0.8231	1,088	6,785	1,690	430	1,736	411
Proprietary Dataset 1	454	54	400	-0.88	0.7547	0.6122	0.8971	0.4615	0.9414	0.8568	30	305	35	19	47	18
Proprietary Dataset 2	507	95	412	-0.81	0.7221	0.6625	0.7817	0.4173	0.9075	0.8264	53	265	74	27	66	22
Proprietary Dataset 3	2,865	171	2,694	-0.94	0.6437	0.4167	0.8707	0.1529	0.9638	0.7899	50	1,866	277	70	500	102
Proprietary Dataset 4	1,278	259	1,019	-0.80	0.6705	0.4952	0.8458	0.4421	0.8716	0.8224	103	713	130	105	151	76
Vitic Intermediates Dataset	1,748	584	1,164	-0.67	0.7541	0.7940	0.7141	0.5754	0.8766	0.8129	370	682	273	96	290	37

Abbreviations

BAC = balanced accuracy: (SEN + SPEC) / 2

SEN = sensitivity: TP / (TP + FN)

SPEC = specificity: TN / (TN + FP)

PPV = positive predictivity: TP / (TP + FP)

NPV = negative predictivity: TN / (TN + FN)

COV = coverage: (TP + FP + TN + FN) / (TP + FP + TN + FN + EQ + OD)

TP = true positive

FP = false positive

TN = true negative

FN = false negative

EQ = equivocal

OD = outside domain

Structure Standardisation

Nexus standardises all input structures before processing them through its applications. The standardisation techniques Sarah uses to convert user-drawn structures into standard forms employs a set of transform rules which include consideration of mixtures. Whereas all components of a query mixture are assessed by Sarah Nexus, pharmaceutically acceptable salts are removed during standardisation, so they are not considered during the prediction. The list of pharmaceutically accepted salts was developed based on Haynes et al, Journal of Pharmaceutical Sciences, 2005, 94, 2111-2120.

Following an assessment of mixture components which are non-mutagens in the Sarah training set, or have published negative Ames result, the list of salts removed by Nexus during standardisation has been updated in Sarah Nexus 3.2. The following table shows the new salts removed in Sarah Nexus 3.2.

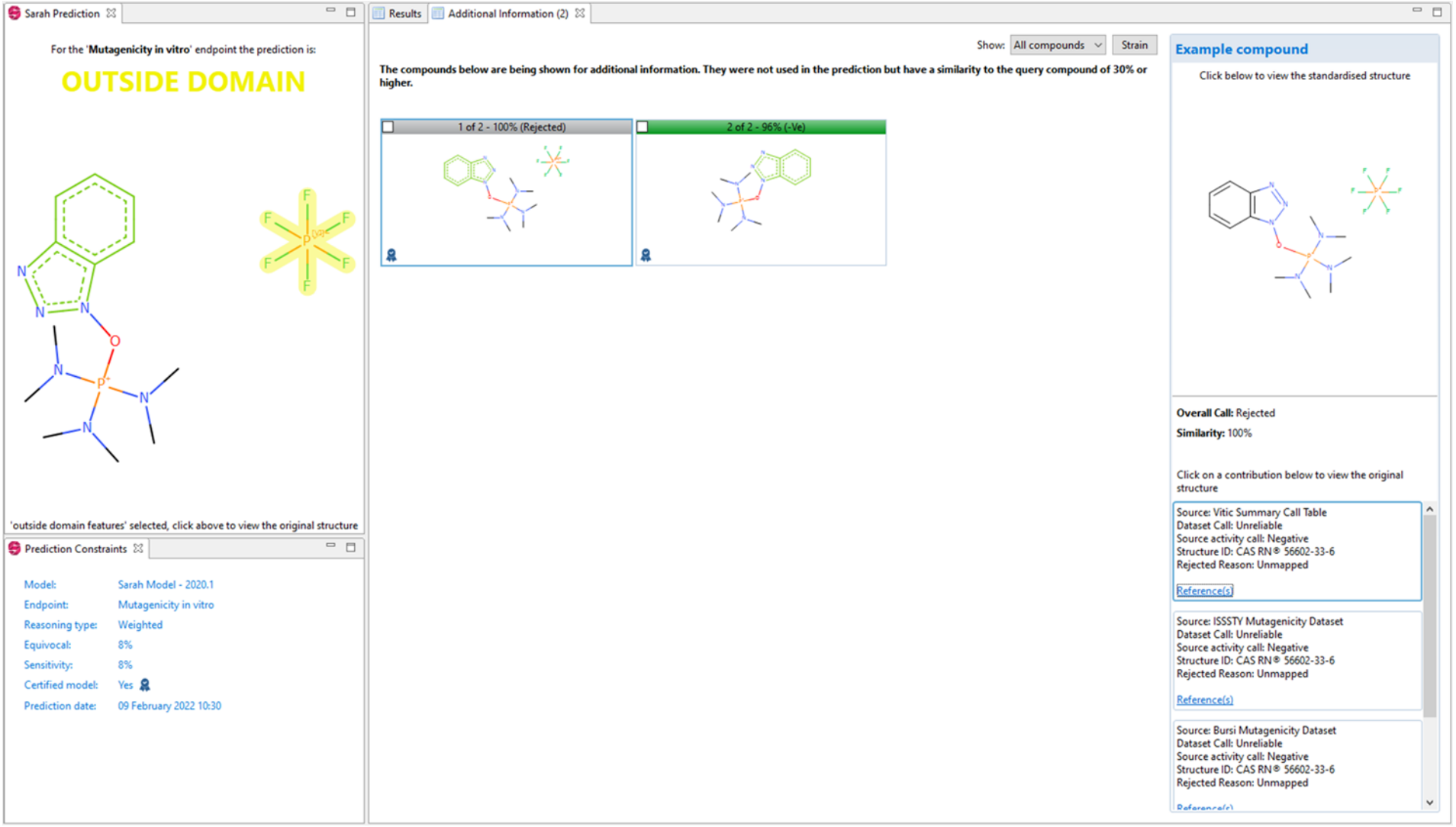
Common Name	SMILES	Reason
Tetrafluoroborate	[B-](F)(F)(F)F	Non-mutagen in training set
Hexafluorophosphate	[P-](F)(F)(F)(F)(F)F	Negative study for lithium hexafluorophosphate*
Ethanolamine	NCCO	Non-mutagen in training set
Triisopropylamine	N(CC(O)C)(CC(O)C)CC(C)O	Non-mutagen in training set
Urea	NC(=O)N	Non-mutagen in training set
Guanidine	NC(=N)N	Non-mutagen in training set
Benzoic acid	C1=CC=CC=C1C(=O)O	Non-mutagen in training set
Ethyl sulfate	O(S([O-])(=O)=O)CC	Non-mutagen in training set
2-Hydroxyethanesulfonic acid	O=S(=O)(CCO)O	Non-mutagen in training set
Salicyclic acid	C1(=CC=CC=C1O)C(=O)O	Non-mutagen in training set
Glycerophosphoric acid	OCC(O)COP(O)(O)=O	Non-mutagen in training set
Beta-Glycerophosphoric acid	P(O)(=O)(OC(CO)CO)O	Non-mutagen in training set

*A negative Ames test has been published for lithium hexafluorophosphate as part of a European Chemicals Agency (ECHA) registration dossier, available at https://echa.europa.eu/registration-dossier/-/registered-dossier/13201/7/7/2.

The following screenshots demonstrate the change in behaviour when processing a query containing a salt that is removed during standardisation, using tri(dimethylamino)benzotriazol-1-yloxyphosphonium hexafluorophosphate (CAS number = 56602-33-6) as an example.

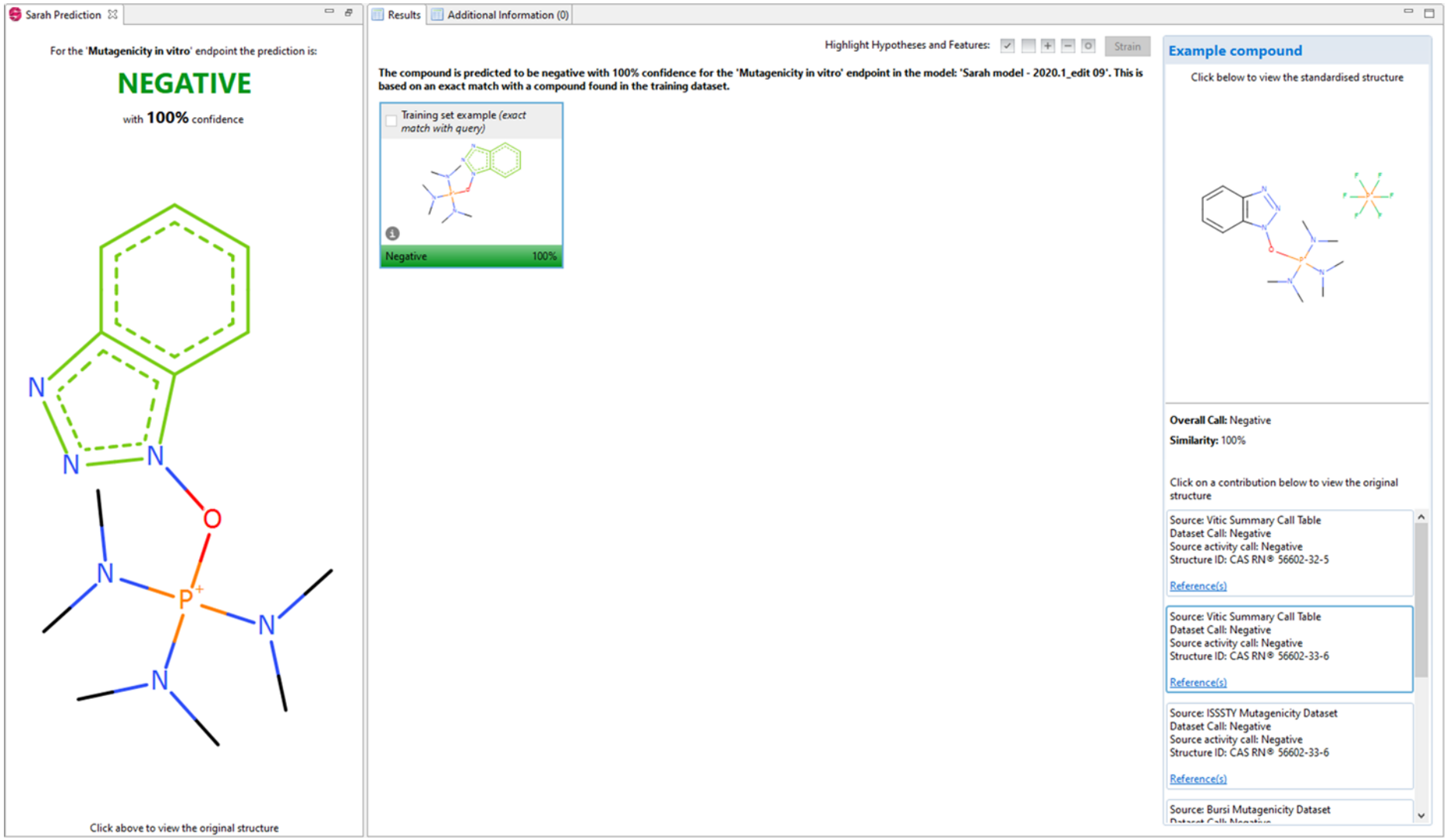
Behaviour in Nexus 2.4, Sarah Nexus 3.1, Sarah Model – 2020.1

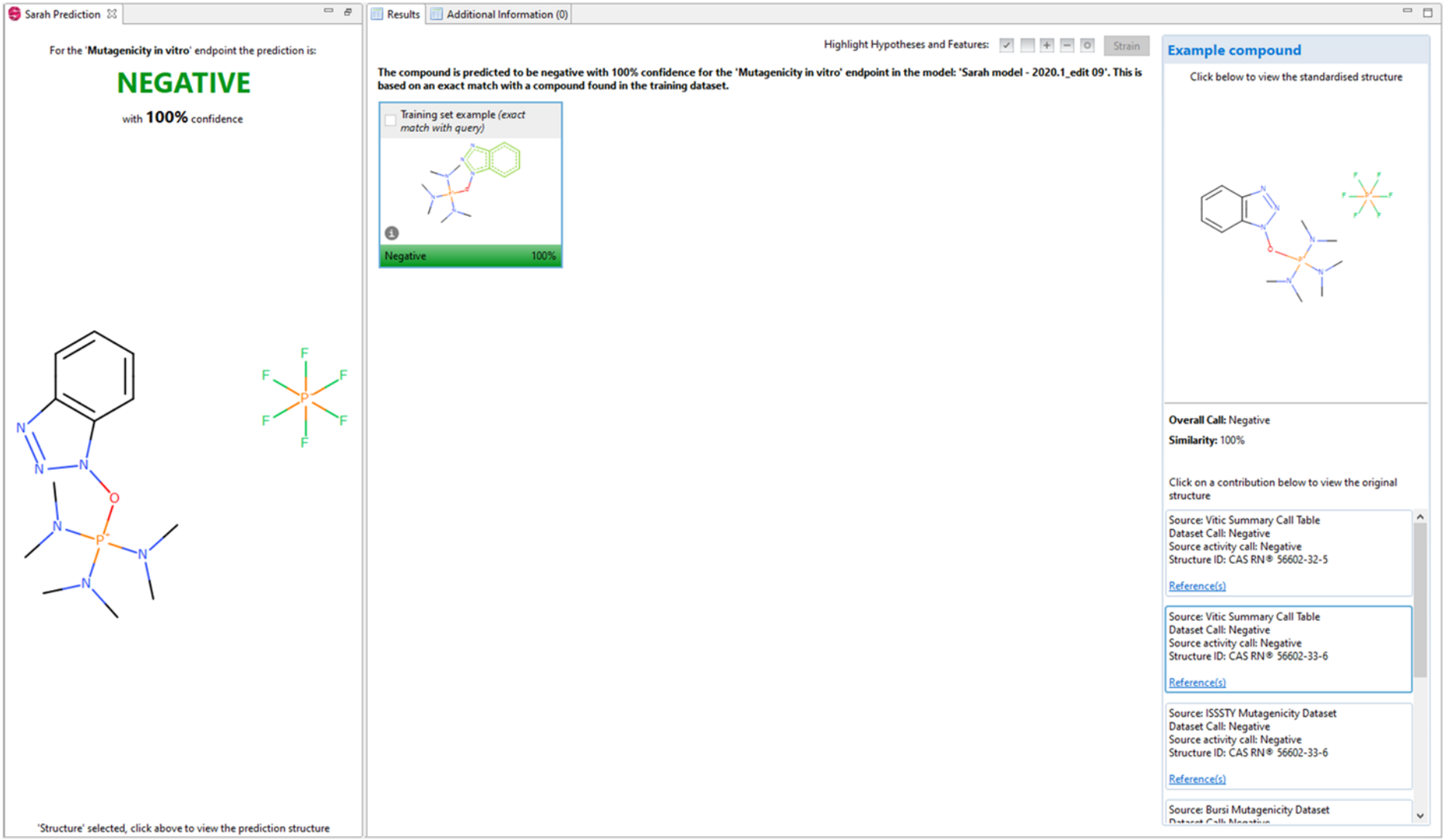
The query is assessed as a mixture to be outside domain, highlighting that the hexafluorophosphate ion is not in the model. The query compound (CAS number = 56602-33-6) is in the additional information tab, as it is in the extended training set but has been rejected by the model as a mixture. The neutral form of the compound (CAS number = 56602-32-5) is also in the additional information tab, as the model treats this as a separate compound.



Behaviour in Nexus 2.5, Sarah Nexus 3.2, Sarah Model – 2022.1

The query is considered an exact match with the compound in the training set as the hexafluorophosphate ion is not assessed. The model considers the salt (CAS number = 56602-33-6) and neutral form (CAS number = 56602-32-5) to be the same; however, clicking on the contribution for each example compound will show the original structure for that training example to assess whether it was the salt. The user may see the original query entered by clicking on the prediction structure.




Sarah Prediction
For the 'Mutagenicity in vitro' endpoint the prediction is:
NEGATIVE
with 100% confidence
'Structure' selected, click above to view the prediction structure
Results
Additional Information (0)
Highlight Hypotheses and Features:
Strain
The compound is predicted to be negative with 100% confidence for the 'Mutagenicity in vitro' endpoint in the model: 'Sarah model - 2020.1_edit 09'. This is based on an exact match with a compound found in the training dataset.
Training set example (exact match with query)
Negative 100%
Example compound
Click below to view the standardised structure
Overall Call: Negative
Similarity: 100%
Click on a contribution below to view the original structure
Source: Vitic Summary Call Table
Dataset Call: Negative
Source activity call: Negative
Structure ID: CAS RN® 56602-32-5
Reference(s)
Source: Vitic Summary Call Table
Dataset Call: Negative
Source activity call: Negative
Structure ID: CAS RN® 56602-33-6
Reference(s)
Source: ISSSTY Mutagenicity Dataset
Dataset Call: Negative
Source activity call: Negative
Structure ID: CAS RN® 56602-33-6
Reference(s)
Source: Bursi Mutagenicity Dataset

ICH M7 Changes in Previous Versions

The following changes were made in previous Nexus 2.5 releases.

ICH M7 Expert Review (Derek 6.2 and Sarah 3.2)

The following changes have been made to the automated expert review arguments displayed following an ICH M7 prediction.

Automated Expert Review Arguments

The list of automated arguments has been updated and renumbered to more clearly group arguments based on the same prediction scenario. New arguments have been implemented, and existing arguments updated, relating to the following categories:

- Chemical classes for which the Ames text may not adequately assess the mutagenic hazard
- Chemicals considered to belong to a chemical class listed in the cohort of concern

Expert Review Argument Numbers

New Number	Old Number
1.1	60
1.2	61
2.1	2
2.2	13
3.1	14
3.2	3
4.1	39
4.2	38
5.1	26
6.1	36
6.2	37

New Number	Old Number
7.1	27
7.2	28
8.1	29
8.2	30
9.1	31
9.2	32
10.1	4
10.2	12
11.1	6
11.2	7
12.1	11
12.2	40
12.3	41
13.1	5
14.1	9
14.2	8
15.1	46
15.2	42
16.1	10
16.2	43
17.1	15

New Number	Old Number
17.2	53
18.1	16
18.2	54
19.1	17
19.2	55
20.1	44
20.2	45
21.1	56
21.2	57
22.1	19
22.2	18
23.1	21
23.2	20
24.1	47
24.2	48
25.1	59
25.2	58
26.1	50
26.2	33
27.1	52
27.2	51

New Number	Old Number
28.1	22
29.1	24
29.2	23
29.3	35
29.4	25
30.1	62
30.2	63
30.3	
30.4	
30.5	
30.6	
30.7	
31.1	1
31.2	
31.3	
31.4	
31.5	

Ames Test Not Adequate

The following arguments represent a query which belongs to a chemical class for which the related Derek alert description discusses reasons why the standard Ames test may not be adequate to assess the mutagenic hazard presented by the chemical class. They have been implemented with an “Inconclusive” overall *in silico* call to notify the user to assess the relevance of the Derek comments for their review.

Argument Number	Chemical Class
31.1	Carboxylic acid halide, carbamoyl halide, sulfonyl halide, thionyl halide
31.2	Allylbenzene
31.3	N-Methylol
31.4	Alkyl aldehyde
31.5	Benzyl halide

Cohort of Concern

The following arguments represent a query which belongs to a chemical class suspected of belonging to a cohort of concern as listed within the ICH M7 guideline, and the user is notified of the requirement to conduct a compound-specific risk assessment. 5 new arguments have been implemented to specifically discuss subclasses of N-nitroso compounds for which structural features are expected to reduce the mutagenic potential as discussed in the related Derek alert 007. These have been implemented with either a “Positive”, “Negative” or “Inconclusive” overall *in silico* call.

Argument Number	Argument Summary	Argument Assignment
30.1	Query belongs to a chemical class considered a cohort of concern; a compound-specific risk assessment has concluded positive.	Positive
30.2	Query belongs to a chemical class considered a cohort of concern; a compound-specific risk assessment has concluded negative.	Negative
30.3	Query is an N-nitroso compound that cannot undergo the expected mechanism that leads to potent activity in other N-nitroso compounds due to a lack of alpha-hydrogen.	Negative
30.4	Query is an N-nitroso compound that cannot undergo the expected mechanism that leads to potent activity in other N-nitroso compounds due to a lack of alpha-hydrogen; however, an alternative mechanism may be possible.	Inconclusive
30.5	Query is an N-nitroso compound that may not be able undergo the expected mechanism that leads to potent activity in other N-nitroso compounds due to being in a hindered cyclic compound.	Negative
30.6	Query is an N-nitroso compound with an alpha-heteroatom substituent, examples of which are mutagenic but have been observed to have weak carcinogenic potency.	Positive
30.7	Query is an N-nitroso compound of a primary amine which is expected to undergo formation of a mutagenic diazonium ion; however, stability in vivo is likely to be sufficiently poor and prevent DNA alkylation.	Positive

Links Between Derek and Sarah

Links between Derek alerts with Sarah hypotheses and training set examples have been updated.

118 Sarah hypotheses have been linked to 55 Derek alerts.

- Sarah hypotheses may be linked to multiple Derek alerts
- Derek alerts may be linked to multiple Sarah hypotheses

6,179 Sarah training examples have been linked to 139 Derek alerts.

- Sarah training examples may be linked to multiple Derek alerts

ICH M7 Classification (Derek 6.2 and Sarah 3.2)

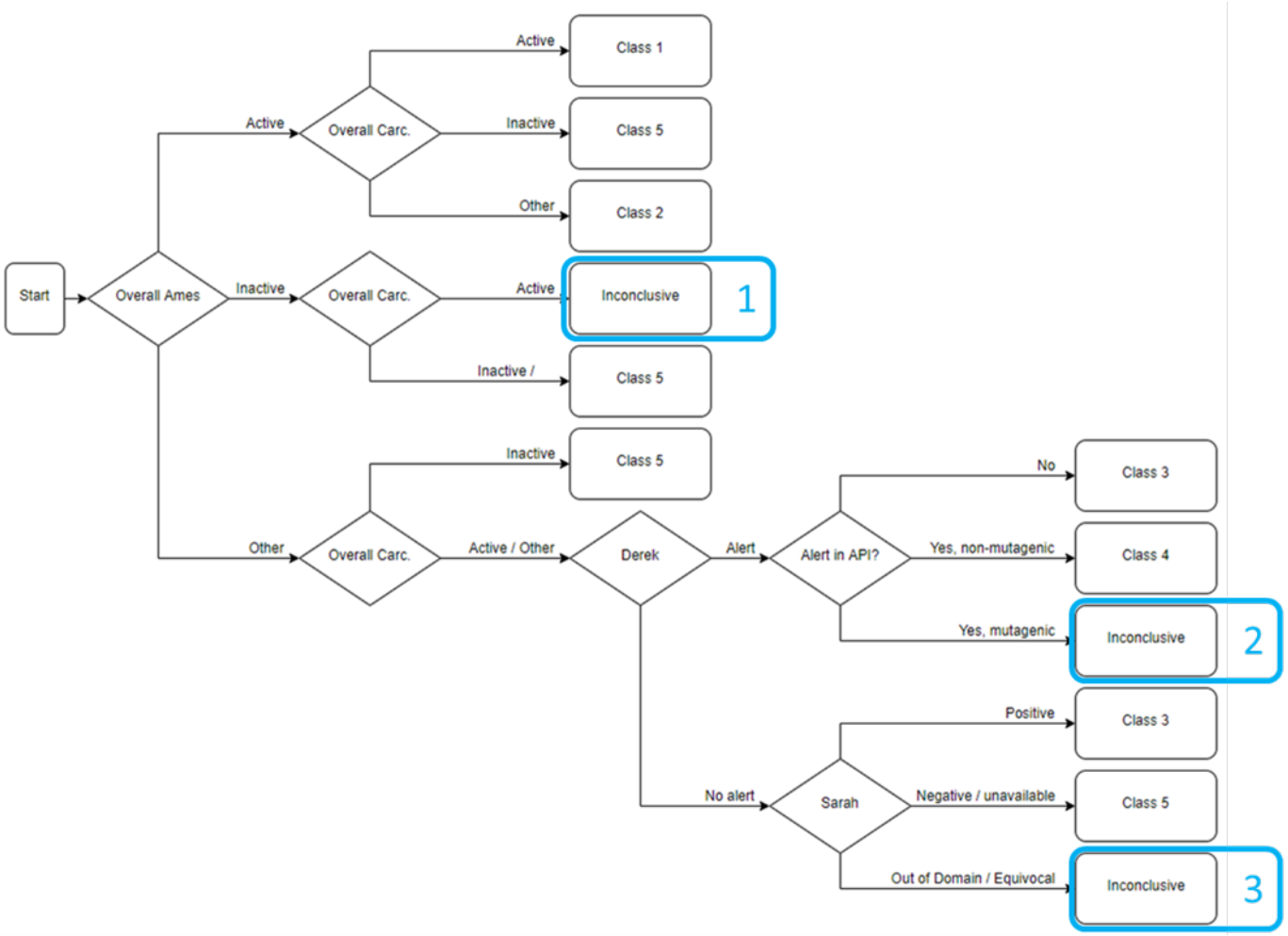
The following changes have been made to the ICH M7 classification tool.

Ames Dataset

The Lhasa Ames dataset has been updated in line with the Sarah training set and the Mutagenicity Negative Predictions dataset.

ICH M7 Class Assignment

A predicted ICH M7 class is displayed in the ICH M7 classification table for queries depending on both the predictions provided by Derek and Sarah as well as consideration of any known Ames or carcinogenicity data. The class is assigned according to the following logic process:



There are 3 scenarios where “Inconclusive” may be assigned instead of ICH M7 class 1 – 5:

1. Overall Ames = inactive and overall carcinogenicity = active
    - Query is a known non-mutagenic carcinogen
2. Overall Ames = other and overall carcinogenicity = active/other and Derek = alert activated by query and API
    - Query is an impurity in a mutagenic API

3. Overall Ames = other and overall carcinogenicity = active/other and Derek = no alert and Sarah = outside domain/equivocal
    - The *in silico* prediction made by Derek and Sarah is considered inconclusive and requires review

Scenario 3 represents the situation where the calculated *in silico* prediction is inconclusive and requires expert review to determine the ICH M7 class for the query. Instead, scenarios 1 and 2 represent situations where the query should be considered outside ICH M7 as the guideline does not apply to drug products intended for advanced cancer indications or drug substances which are genotoxic at therapeutic concentrations. The workflow has been updated to better represent this, showing “Inconclusive (Outside scope of ICH M7)” for scenarios 1 and 2.

Permissible Daily Exposure Comments

Comments are displayed in the ICH M7 classification table for queries which have had a permissible daily exposure (PDE) or acceptable intake (AI) published. These comments are displayed for any component of the query that matches a chemical in the published list of PDEs and AIs.

The following updates have been made to the information presented:

- ICH guideline M7 – addendum
    - Entries updated in line with changes in updated publication October 2021
- ICH guideline Q3C
    - Entries updated in line with changes in updated publication (R8) May 2021
- Literature publications
    - New entries added for 5 chemicals listed in the following table

Chemical	CAS	Limit	Reference
Diisopropyl ether	108-20-3	PDE = 980 ug/day (inhalation)	Romanelli and Evandri, Toxicological Research, 2018, 34, 111-125
Irganox 1010	6683-19-8	PDE = 8,000 ug/day (parenteral)	Parris et al, Regulatory Toxicology and Pharmacology, 2020, 118, 104802
Irgafos 168	31570-04-4	PDE = 2,900 ug/day (parenteral)	Parris et al, Regulatory Toxicology and Pharmacology, 2020, 118, 104802
Bisphenol A	80-05-7	PDE = 4.2 ug/day (parenteral)	Parris et al, Regulatory Toxicology and Pharmacology, 2020, 118, 104802
Butylated hydroxytoluene	128-37-0	PDE = 12,500 ug/day (parenteral)	Parris et al, Regulatory Toxicology and Pharmacology, 2020, 118, 104802

Our Products

Derek® nexus — *Expert knowledge-based toxicity prediction software.*

Sarah® nexus — *Statistical-based software for the prediction of mutagenicity.*

Mirabilis® — *A tool for assessing the relative purging of mutagenic impurities.*

Zeneth® — *Expert decision support software for predicting the forced degradation pathways of organic compounds.*

Vitic® — *The chemical database and information management system, offering researchers and scientists rapid access to searchable toxicological information.*

Meteor® nexus — *Expert decision support software for predicting the metabolic fate of chemicals in mammals.*

Effiris™ — *A secondary pharmacology model suite leveraging value from federated learning.*

Kaptis™ — *A tool to support risk assessment in the context of adverse outcome pathways.*

Lhasa Limited

shared **knowledge** • shared **progress**

Lhasa Limited Registered Office
Granary Wharf House, 2 Canal Wharf, Leeds LS11 5PS
Registered Charity (290866)

+44 (0)113 394 6020
info@lhasalimited.org
www.lhasalimited.org