

## Use lazar-GUI for predictions with QSAR models based on the lazar-OpenTox framework

David Vorgrimmler and Micha Rautenberg, *in silico* toxicology Basel, Switzerland

### Introduction

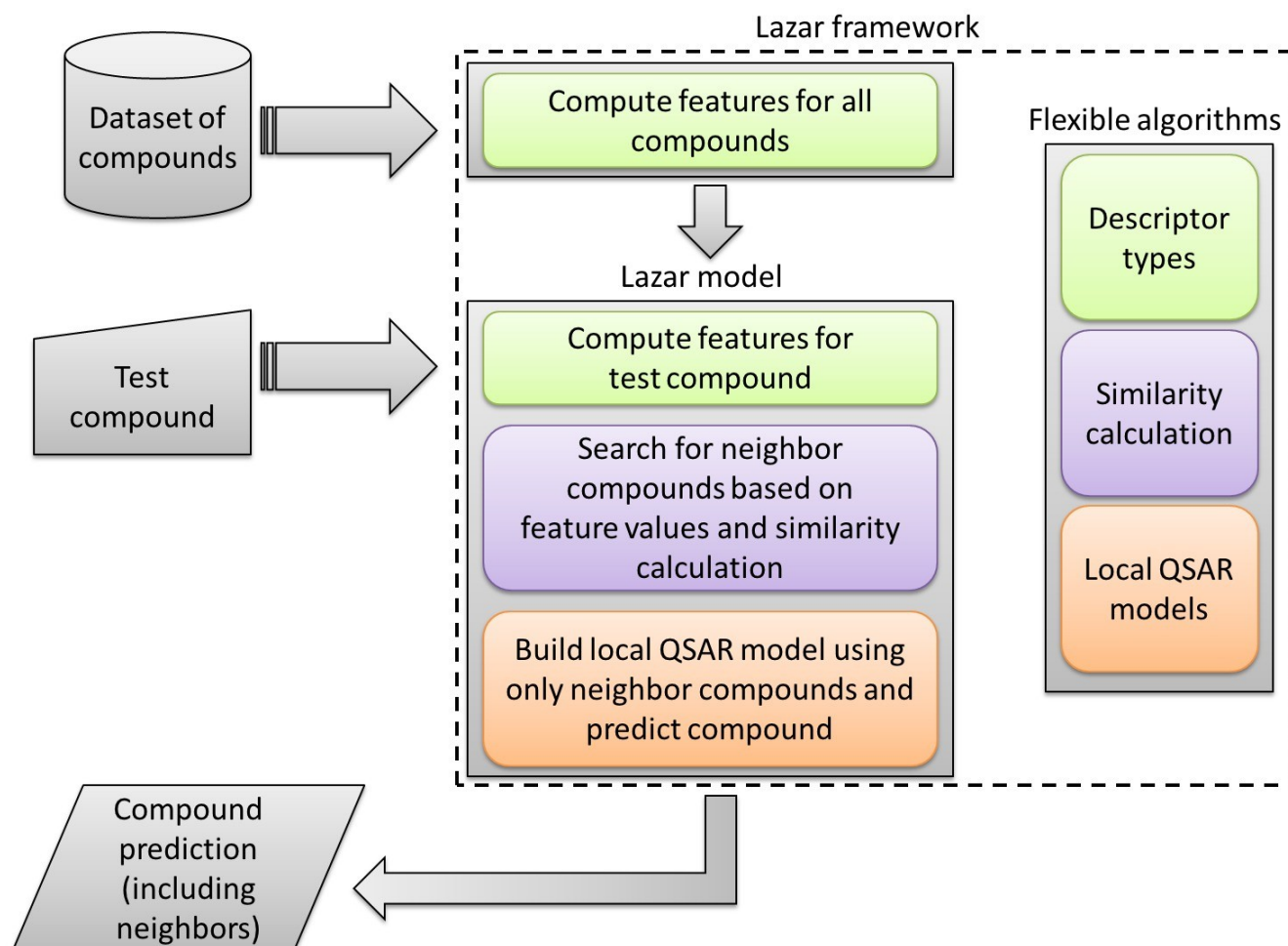
*lazar*<sup>1</sup> (lazy structure-activity relationships) is a modular framework for predictive toxicology. Similar to the read across procedure in toxicological risk assessment, *lazar* creates local QSAR (quantitative structure-activity relationship) models for each compound to be predicted. Model developers can choose between a large variety of algorithms for descriptor calculation and selection, chemical similarity indices for model building. The *lazar*-GUI is a graphical webinterface for the *lazar* framework, which enables user to make predictions on existing QSAR models.

In this workshop participants will use the *lazar*-GUI to make predictions for chemical compounds on classification and regression models. The available models are based on the *lazar*-OpenTox framework.

### Full compatibility with OpenTox

Each component and algorithm of the *lazar* framework is implemented as a OpenTox<sup>2</sup> web service. This makes our framework modular, each algorithm can be easily replaced. Each web service can be used standalone (e.g. algorithm, dataset, model, ...). The source code of the framework is available as open-source on GitHub (<https://github.com/opentox>). The *lazar*-GUI is located at <https://services.in-silico.ch/predict>.

### Workflow



## ***Descriptor types***

### *Substructure fragments :*

- Substructure descriptors are binary descriptors that do occur in a compound or not.
- *lazar* integrates algorithms to mine endpoint relevant substructures.
  - Backbone Refinement Class Mining (BBRC)
  - Latent Structure Pattern Mining (LAST)

### *Physico-chemical (PC) properties*

- Numerical descriptors (e.g. molecular weight, largest chain, logP, surface charge).
- *lazar* utilizes open source cheminformatics libraries (Chemistry Development Kit (CDK), OpenBabel, JOELib) to compute PC descriptors.

Also Measured properties (experimental measurements) can be utilized by *lazar*.

## ***Similarity calculation***

### *Substructures*

- A weighted Tanimoto index searches for similar compounds (neighbors) to the query compound in the training dataset .
- Tanimoto similarity encodes presence or absence of substructures in molecules
- Optionally, the number of times a substructure occurs in each compound is taken into account

### *Physico-chemical properties / numerical features*

- Cosine similarity, based on the PCA-transformed feature space, is used to find the neighbors
- Cosine similarity computes the distance between two compounds by measuring the angle between the feature value vectors

## ***Local QSAR models***

Only neighbor compounds are used to build a local QSAR model.

### *Available learning algorithms*

- Classification: weighted majority voting and support vector machines (SVM)
- Regression: support vector machines (SVM) with a variety of kernels

### *Applicability Domain*

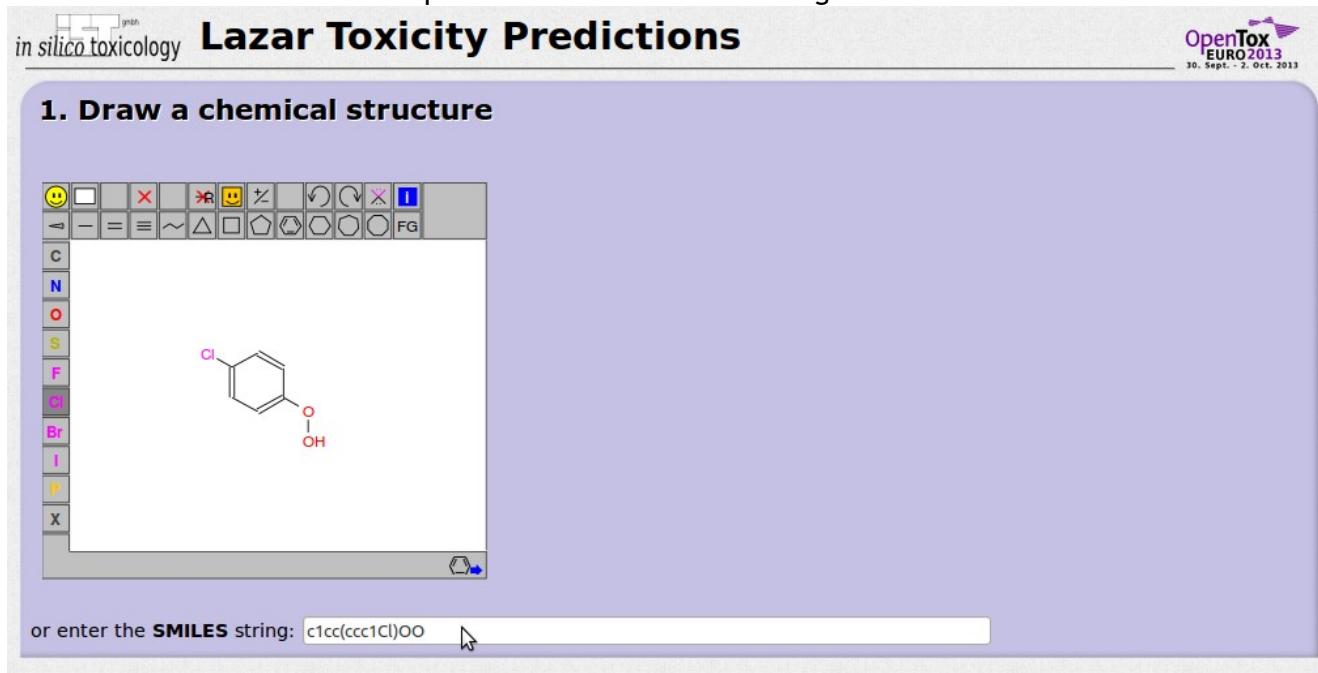
- *lazar* provides a confidence value with each prediction, that measures how
- ‘certain’ the algorithm is about the prediction
- The confidence is based on:
  - The number of compounds (neighbors) that are similar to the query structure
  - The degree of similarity of the neighbors to the query compound
  - The coherence of experimental data within neighbors.

## Hands-on Workshop Session

**Requirements:** Browser with Javascript activated (we recommend Firefox), Java installed, Internet connection.

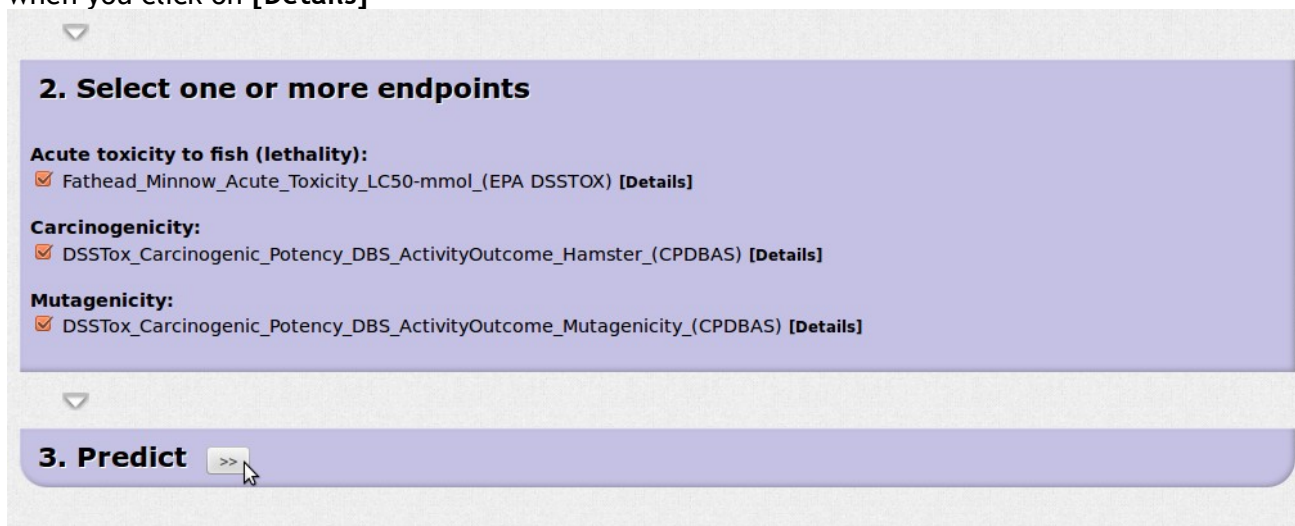
### Part 1 - Make a prediction

Draw a chemical structure/compound<sup>3</sup> or enter a Smiles string.



The screenshot shows the 'Lazar Toxicity Predictions' web application. At the top left is the 'in silico toxicology' logo. The title 'Lazar Toxicity Predictions' is in the center. On the right is the 'OpenTox EURO2013' logo with the dates '30. Sept. - 2. Oct. 2013'. Below the title is a section '1. Draw a chemical structure'. It contains a chemical drawing toolbar with various icons for drawing rings, bonds, and atoms. Below the toolbar is a canvas showing a chemical structure of 4-chlorobenzoic acid. At the bottom of this section is a text input field labeled 'or enter the SMILES string:' with the value 'c1cc(ccc1Cl)OO' entered.

Selected one or more endpoints. You find additional informations and a link to validation-results when you click on **[Details]**



The screenshot shows the '2. Select one or more endpoints' section of the web application. It lists three categories of endpoints, each with a checked checkbox and a '[Details]' link:

- Acute toxicity to fish (lethality):**
  - ☒ Fathead\_Minnow\_Acute\_Toxicity\_LC50-mmol\_(EPA\_DSSTOX) [\[Details\]](#)
- Carcinogenicity:**
  - ☒ DSSTox\_Carcinogenic\_Potency\_DBS\_ActivityOutcome\_Hamster\_(CPDBAS) [\[Details\]](#)
- Mutagenicity:**
  - ☒ DSSTox\_Carcinogenic\_Potency\_DBS\_ActivityOutcome\_Mutagenicity\_(CPDBAS) [\[Details\]](#)

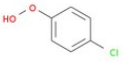
Below this section is a section '3. Predict' with a button labeled '>>'.

Run a prediction, click on "Predict" and wait for the result page.

**Lazar Toxicity Predictions**

**New Prediction**

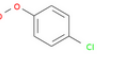
**Prediction Results:**

 <p><b>Names and synonyms</b></p>	<b>Fathead Minnow Acute Toxicity LC50-mmol (EPA DSSTOX)</b> <b>Type:</b> regression <b>Result:</b> 0.784 ⓘ <b>Confidence:</b> 0.782 ⓘ <b>Descriptors</b> <b>Neighbors:</b> ▼	<b>DSSTox Carcinogenic Potency DBS ActivityOutcome Hamster (CPDBAS)</b> <b>Type:</b> classification <b>Result:</b> inactive ⓘ <b>Confidence:</b> 0.411 ⓘ <b>Significant fragments</b> <b>Neighbors:</b> ▼	<b>DSSTox Carcinogenic Potency DBS ActivityOutcome Mutagenicity (CPDBAS)</b> <b>Type:</b> classification <b>Result:</b> inactive ⓘ <b>Confidence:</b> 0.283 ⓘ <b>Significant fragments</b> <b>Neighbors:</b> ▼

© in-silico toxicology gmbh 2004-2013

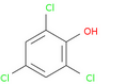
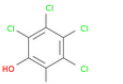
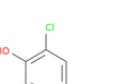
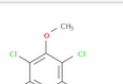
Click on **Neighbors** to examine the predicted compound's neighbors with relevant substructures, measured activity and similarity.

**Prediction Results:**

 <p><b>Names and synonyms</b></p>	<b>Fathead Minnow Acute Toxicity LC50-mmol (EPA DSSTOX)</b> <b>Type:</b> regression <b>Result:</b> 0.784 ⓘ <b>Confidence:</b> 0.782 ⓘ <b>Descriptors</b> <b>Neighbors:</b> ▼	<b>DSSTox Carcinogenic Potency DBS ActivityOutcome Hamster (CPDBAS)</b> <b>Type:</b> classification <b>Result:</b> inactive ⓘ <b>Confidence:</b> 0.411 ⓘ <b>Significant fragments</b> <b>Neighbors:</b> ▼	<b>DSSTox Carcinogenic Potency DBS ActivityOutcome Mutagenicity (CPDBAS)</b> <b>Type:</b> classification <b>Result:</b> inactive ⓘ <b>Confidence:</b> 0.283 ⓘ <b>Significant fragments</b> <b>Neighbors:</b> ▼

LC50-mmol Hamster **Mutagenicity**

**Neighbors:**

Compound	Measured Activity ⓘ	Similarity ⓘ	Supporting information
	inactive	1.0	Names and synonyms Significant fragments
	inactive	1.0	Names and synonyms Significant fragments
	inactive	1.0	Names and synonyms Significant fragments
	active	0.938	Names and synonyms Significant fragments

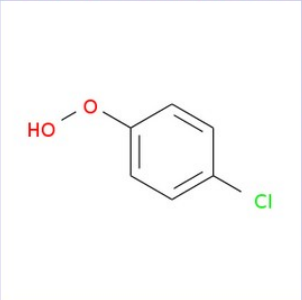


Click on **Descriptors** for regression or **Significant fragments** for classification models to inspect the given features of the compound.

**Descriptors** shows the description of the physico-chemical properties and its values.

**Significant fragments** shows structural features sorted by their effects (see vocabulary). Each effect list is sorted by the p value of the substructure which was calculated during the feature generation. The substructures are written in SMARTS notation<sup>4</sup>.

**Significant fragments:**

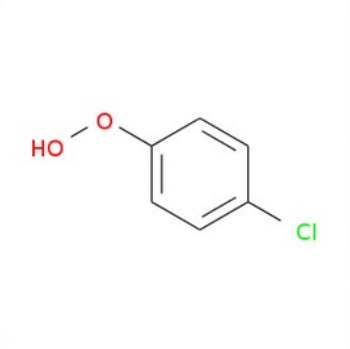


Predominantly in compounds with activity "inactive"	p value
[ #17&A]-[ #6&a]	0.99907
[ #8&A]-[ #6&a]	0.99692
[ #17&A]-[ #6&a]:[ #6&a]	0.99825
[ #17&A]-[ #6&a]:[ #6&a]:[ #6&a]	0.99825
[ #8&A]-[ #6&a]:[ #6&a]:[ #6&a]	0.99773
[ #17&A]-[ #6&a]:[ #6&a]:[ #6&a]:[ #6&a]	0.99825
[ #17&A]-[ #6&a]:[ #6&a]:[ #6&a]:[ #6&a]-[ #8&A]	0.99697
[ #17&A]-[ #6&a]:[ #6&a]:[ #6&a]:[ #6&a]:[ #6&a]	0.99825
[ #8&A]-[ #6&a]:[ #6&a]:[ #6&a]:[ #6&a]:[ #6&a]	0.9955
[ #17&A]-[ #6&a]:[ #6&a]:[ #6&a]:[ #6&a]:[ #6&a]:[ #6&a]	0.99674
[ #17&A]-[ #6&a]:([ #6&a]:[ #6&a])(:[ #6&a]:[ #6&a]:[ #6&a]-[ #8&A])	0.99519
[ #17&A]-[ #6&a]:([ #6&a]:[ #6&a])(:[ #6&a]:[ #6&a]:[ #6&a])	0.99557
[ #8&A]-[ #6&a]:([ #6&a]:[ #6&a])(:[ #6&a]:[ #6&a]:[ #6&a])	0.99445
[ #17&A]-[ #6&a]:([ #6&a]:[ #6&a])(:[ #6&a]:[ #6&a])	0.99557
[ #8&A]-[ #6&a]:([ #6&a]:[ #6&a])(:[ #6&a]:[ #6&a])	0.99445

Predominantly in compounds with activity "active"	p value
---	---------

Click on **Names and synonyms** get following informations.

**Names and synonyms:**



**SMILES:**  
c1cc(ccc1Cl)OO

**InChI:**  
1S/C6H5ClO2/c7-5-1-3-6(9-8)4-2-5/h1-4,8H

**Names:**  
1-chloro-4-hydroperoxybenzene; 1-chloro-4-hydroperoxy-benzene


**PubChem read across (experimental)**

## Part 2 - Make a prediction and a PubChem read across investigation

Click on **New Prediction**


Draw a chemical structure or enter a Smiles string e.g. "c1ccc(cc1)O" (PHENOL).

Selected one or more endpoints and start a prediction.



in-silico toxicology

Lazar Toxicity Predictions




OpenTox

EURO2013

30. Sept. - 2. Oct. 2013

New Prediction



Names and synonyms

Fathead Minnow Acute Toxicity LC50-mmol (EPA DSSTOX)

Type: regression

Database hit:  
Compound found in training dataset

Measured activity: 0.347

DSSTox Carcinogenic Potency DBS ActivityOutcome Hamster (CPDBAS)

Type: classification

Result: inactive ⓘ

Confidence: 0.411 ⓘ

Significant fragments

Neighbors: ⚡

DSSTox Carcinogenic Potency DBS ActivityOutcome Mutagenicity (CPDBAS)

Type: classification

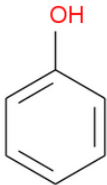
Database hit:  
Compound found in training dataset

Measured activity: inactive

© in-silico toxicology gmbh 2004-2013

There are two database hits for this compound.

Click on **Names and synonyms** and copy “PHENOL” for the PubChem read across search.

<p><b>Names and synonyms:</b></p> 	<p><b>SMILES:</b></p> <p>c1ccc(cc1)O</p> <p><b>InChI:</b></p> <p>1S/C6H6O/c7-6-4-2-1-3-5-6/h1-5,7H</p> <p><b>Names:</b></p> <p><b>PHENOL</b>; 1336-35-2; 108-95-2; 63496-48-0; 73607-76-8; 61788-41-8; 14534-23-7; 50356-25-7; 8002-07-1; 27073-41-2; AIDS000352; Phenol homopolymer; Phenosmolin; TEA polyphenol; P6471_SIGMA; Benzenol; C00146; Hydroxybenzene; Phenic acid; Phenylic acid; Carbolsaure; D06535; Liquefied phenol; Liquefied phenol (JP15); Liquefied phenol (TN); Phenol, liquefied (USP); D06536; Phenol for disinfection; Phenol for disinfection (JP15); Phenol for disinfection (TN); Phenol, liquefied (USP); D06536; Phenol for disinfection; Phenol for disinfection (JP15); Phenol for disinfection (TN);</p> <p>Benzene, hydroxy-; CCRIS 504; Liquefied Phenol; P9346_SIAL; Baker's p and s; A13-01814; Acide carbolique [French]; Phenol, chlorinated; Phenol, chloro derivs.; Carbollic oil; Carbolsaure [German]; Caswell No. 649; EINECS 203-632-7; EPA Pesticide Chemical Code 064001; FEMA No. 3223; Fenol [Dutch, Polish]; Fenolo [Italian]; HSDB 113; Liquid phenol; NSC 36808; Phenol (or solutions with 5% or more phenol); Phenol [JAN]; Phenol alcohol; Phenol solutions [UN2821] [Poison]; Phenol, molten [UN2312] [Poison]; Phenol, pure; Phenol, solid [UN1671] [Poison]; Phenole [German]; NCGC00091454-01; RCRA waste number U188; UN 1671 (solid); UN 2312 (molten); UN1671; UN2312; UN2821; Un 2812 (solution); 185450_SIAL; 48556_SUPELCO; 48688_SUPELCO; c0128; 328111_SIAL; Phenic; P3653_SIAL; Monohydroxy benzene; CHEBI:15882; 46344_RIEDEL; Karbolsaure; acide phenique; P1037_SIAL; carbollic acid; phenylalcohol; (14C)Phenol; Phenol, labeled with carbon-14; ZINC00895069; IPH; NCGC00091454-02; NCGC00091454-03; 35952_RIEDEL; D00033; Paoscle (TN); Phenol (JP15/USP); Phenol (TN); P4682_SIAL; 40063_SUPELCO; Phenol solution; Acide carbolique; Baker's P &amp; S liquid &amp; Ointment; Baker's P and S Liquid and Ointment; Carbollic acid, liquid; Fenol; Fenolo; Izal; Monohydroxybenzene; Monophenol; NCI-C50124; NSC36808; Oxybenzene; Paoscle; PhOH; Phenol (liquid); Phenol, liquefied; Phenol, liquid; Phenol, synthetic; Phenole; Phenyl alcohol; Phenyl hydrate; Phenyl hydroxide; Phenylic alcohol; Synthetic phenol; WLN: QR; P4557_SIAL; W322318_ALDRICH; 52463_FLUKA; 297437_SIGMA; Phenol-UL-14C; C15584; RCRA waste no. U188; Fenosmolin; 564796_ALDRICH; Phenol, polymer-bound; 33517_RIEDEL; 16016_RIEDEL; 77607_FLUKA; 16017_RIEDEL; 77614_FLUKA; 16018_RIEDEL; P5566_SIAL; EINECS 262-972-4; Phenol, sulfurated; ST5214353; InChI=1/C6H6O/c7-6-4-2-1-3-5-6/h1-5,7; LS-476; AIDS-000352; Fenosmoline; 242322_SIAL</p> <p><b>PubChem</b> read across (experimental)</p> <p>link opens in new tab</p>
---	---

Click on **PubChem read across** to open a new tab with our service. Enter and search for “PHENOL”. In your investigation you will find a lot of information for this compound including the underlying assays of our models.

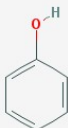
Please keep in mind that our PubChem read across service is still experimental.

### PubChem read across

Compound name:

This is an experimental version. Loading data from PubChem can be slow. Please use the "Back" button and retry the offending operation if you have timeout problems.

**phenol CID: 996**



Gene/protein targets (experimental)

Gene/protein targets (read across)

Gene/protein non-targets (experimental)

Gene/protein non-targets (read across)

autoinducer synthetase family protein [Burkholderia mallei ATCC 23344]  
Target GeneID: 52423132  
Assay ID: 720554  
p\_active: 0.028  
p\_inactive: 0.972

replicative DNA helicase [Staphylococcus aureus subsp. aureus JH1]  
Target GeneID: 150392496  
Assay ID: 485395  
p\_active: 0.028  
p\_inactive: 0.972

67.9K protein [Vaccinia virus]  
Target GeneID: 222762  
Assay ID: 720580  
p\_active: 0.028  
p\_inactive: 0.972

67.9K protein [Vaccinia virus]  
Target GeneID: 222762  
Assay ID: 720579  
p\_active: 0.028  
p\_inactive: 0.972

Parkin [Homo sapiens]  
Target GeneID: 3063388  
Assay ID: 720573  
p\_active: 0.028  
p\_inactive: 0.972

Parkin [Homo sapiens]  
Target GeneID: 3063388  
Assay ID: 720572  
p\_active: 0.028  
p\_inactive: 0.972

Amyloid beta A4 protein  
Target GeneID: 112927  
Assay ID: 720559

Other active assays (experimental)

Agonist activity at GABAA assessed as enhancement of channel current  
Assay ID: 447577

Apparent Michaelis constant (Km) against Arylsulfotransferase (AST IV)  
Assay ID: 39219

Cytotoxicity against human MIA PaCa2 cells after 4 hrs by MTT assay  
Assay ID: 303043

Cytotoxicity against human MIA PaCa2 cells after 72 hrs by MTT assay  
Assay ID: 303044

DSSTox (EPA/FHM) EPA Fathead Minnow Acute Toxicity  
Assay ID: 1188

Inhibition of Cryptococcus neoformans recombinant Can2 beta-carbonic anhydrase after 15 mins by stopped flow CO2 hydration assay  
Assay ID: 588185

Other active assays (read across)

Cytotoxicity against human MIA PaCa2 cells after 4 hrs by MTT assay  
Assay ID: 303043  
p\_active: 0.979  
p\_inactive: 0.021

Cytotoxicity against human MIA PaCa2 cells after 72 hrs by MTT assay  
Assay ID: 303044  
p\_active: 0.979  
p\_inactive: 0.021

MultiTox-Fluor Cytotoxicity Assay - LYMP1-003 - Live Cells  
Assay ID: 962  
p\_active: 0.972  
p\_inactive: 0.028

Depigmentation activity in mouse Melan-a cells after 4 days by spectrophotometry  
Assay ID: 662540  
p\_active: 0.972  
p\_inactive: 0.028

Inhibition of mushroom tyrosinase using L-tyrosine as substrate after 60 mins by spectrophotometry

## Vocabulary

- **Classification** - Prediction of qualitative properties, e.g. to distinguish between toxic and non-toxic compounds
- **Confidence** - Indicates the applicability domain of a model. Predictions with a high confidence can be expected to be more reliable than predictions with low confidence. Confidence values may take any value between 0 and 1. For most models confidence > 0.025 is a sensible (hard) cutoff to distinguish between reliable and unreliable predictions.
- **InChI** - International Chemical Identifier
- **Measured Activity** - Experimental result(s) from the training dataset.
- **Regression** - Prediction of quantitative properties, e.g. LC50 values
- **Result** - *lazar* calculates searches the training dataset for similar compounds (neighbors) and calculates the prediction from their measured activities. *lazar* calculates predictions using
  - a majority vote (weighted by compound similarity) for classification ( original publication )
  - a local QSAR model based on neighbors for regression ( original publication )Please keep in mind that predictions are based on the measured activities of neighbors.
- **Significant fragments** - Substructures that occur (statistically significant) more frequently in one of the toxicity classes or correlate with quantitative toxicity values. Substructures can take any shape (without cycles) and are determined with the fminer algorithm.
- **Similarity** - *lazar* calculates activity specific similarities based on the presence of statistically significant fragments.  
This procedure will
  - consider only those parts of a chemical structure that are relevant for a particular endpoint
  - ignore inert parts of the structure
  - lead to different similarities, depending on the toxic endpoint.  
Similarities of 1 may be encountered even for structurally dissimilar compounds, because inert parts are ignored.
- **SMILES** - Simplified molecular-input line-entry system

## Literature

- 1. Maunz A., Gütlein M., Rautenberg M., Vorgrimmler D., Gebele D., Helma C.: Lazar: A Modular Predictive Toxicology Framework. *Frontiers in Pharmacology* 4(38), April 2013.  
[http://www.frontiersin.org/predictive\\_toxicity/10.3389/fphar.2013.00038/abstract](http://www.frontiersin.org/predictive_toxicity/10.3389/fphar.2013.00038/abstract)
- 2. Hardy, B., Douglas, N., Helma, C., Rautenberg, M., Jeliaskova, N., et al. (2010). Collaborative development of predictive toxicology applications. *J. Cheminform.* 2, 7.
- 3. JSME Molecule Editor by Peter Ertl and Bruno Bienfait <http://peter-ertl.com/jsme/>
- 4. SMARTS - A Language for Describing Molecular Patterns  
<http://www.daylight.com/dayhtml/doc/theory/theory.smarts.html>

## About us

*in silico* toxicology is a small research oriented company focusing on the development of open source predictive toxicology software and services. We are specialized in developing and applying advanced data mining algorithms for complex chemical and biological data.

## Contact

*in silico* toxicology gmbh  
Burgfelder Str. 190  
CH-4055 Basel, Switzerland

Email: [support@in-silico.ch](mailto:support@in-silico.ch)