

Continuing the public benefit of CPDB

Dr Nicholas Marchetti

Product Manager

nik.marchetti@lhasalimited.org



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

Who are Lhasa Limited?

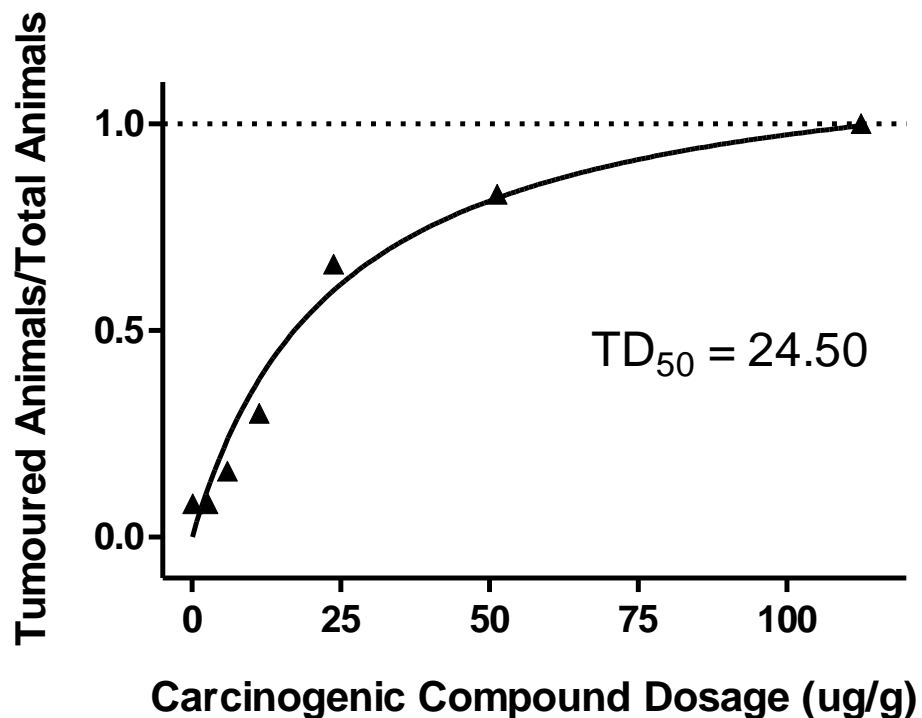
- ▶ Established in 1983
- ▶ Not-for-profit organisation
- ▶ Educational Charity
- ▶ Controlled by our members
- ▶ Knowledge base, Statistical and Database systems



What is CPDB

The Carcinogenic Potency Project

- Carcinogenic Potency DataBase
- Database for carcinogenic effect and calculated TD_{50} (tumour dosage) values of over 1500 compounds



Carcinogenic Compound Dosage (ug/g)	Tumoured Animals (N)	Total Animals (N)	Tumoured Animals/Total Animals
0.0	1	13	0.08
2.6	1	13	0.08
5.9	2	12	0.16
11.2	4	13	0.30
23.7	8	12	0.66
51.2	10	12	0.83
112.4	12	12	1.00

What is CPDB

The Carcinogenic
Potency Project

- Carcinogenic Potency DataBase
- Database for carcinogenic effect and calculated TD₅₀ (tumour dosage) values of over 1500 compounds
 - Created to aggregate the results of different publications and groups under the same umbrella
 - Provides a direct comparison between the studies and a subjective quality score
 - Calculated individual studies TD₅₀, as well as general compound one
- Manually curated for over 30 years by Lois Gold (UC Berkeley)
- Seen as a great resource by the community
- No longer supported – last update in 2007

Where CPDB excels

- CPDB offered an easy, comparable and unbiased way to compare different studies on the same compound
 - **Easy:** searching for the CAS number or the compound name showed all carcinogenicity studies linked to it
 - **Comparable:** the database contained an array of information that made it easy to compare the different studies, such as animal model, lineage, mode of injection, injection area, type of tumour and more
 - **Unbiased:** because they were not content-producers, the commitment was to provide as much information as possible and not value one study over the other without evidence that supported that
- Development of a good method that allows TD_{50} calculation (with some caveats), accounts for spontaneous tumour generation and takes into consideration the length of the study

Where CPDB falls short

Subjectivity & Curation


- Studies were assigned a quality score
- Unclear how this affects TD_{50} calculation

Reproducibility

- TD_{50} values are often not reproducible without eliminating certain datapoints
- Methodology not clear on how data quality affects TD_{50}



Introducing: Lhasa CarcDB

- We want to expand upon CPDB
 - **By maintaining CPDB great features**
 - Public database
 - Up-to-date
 - Individual study annotation (animal model, lineage, tumour type...)
 - Calculated TD₅₀
 - **By improving on what is available**
 - New transparent TD₅₀ calculation method, heavily based on Gold's one, but objective and data-driven (paper coming out soon!)
 - **By adding new features**
 - Structure-based search, on the top of the standard search by CAS number or common name
- 

Why use Lhasa CarcDB?

- Lhasa is a non-profit organization acting as an unbiased curator
- CarcDB is open and free for all
- We are committing to keep CarcDB updated and supported
 - Imported CPDB and expanding with new studies
 - Currently over 1500 compounds and 6000 studies
- No subjective data analysis
- Reproducible TD₅₀ values
- Structure searchable
- Transparent and accessible TD₅₀ methodology
 - Gold's TD₅₀ are shown next to Lhasa's one so the user has as much information as possible

How is Lhasa CarcDB TD₅₀ calculated?

- Based on Gold's method¹
- Controls for natural tumour occurrence and study length
- Use L-BFGS-B optimization algorithm to identify global *minima* parameters, on 3 or more datapoints
- Subjective data analysis no longer necessary
- Compound TD₅₀ is calculated by using a harmonic average on individual studies' TD₅₀
 - Harmonic average: ideal in samples with large diversity, while being a conservative average

¹Peto *et al.*, 1984 and Sawyer *et al.*, 1984

CARCINOGENICITY DATABASE



The Lhasa Carcinogenicity Database is a searchable repository of 6529 long-term carcinogenicity studies covering a total of 1529 chemicals. The database builds upon the work done between 1980 and 2005 by Lois Swirsky Gold and her team. For further details, please see <https://toxnet.nlm.nih.gov/cpdb>.

Some compounds from Gold *et al.* dataset have been combined for the Lhasa Carcinogenicity Database as they represent the same chemical entity (for example at technical grade and commercial grade).

[More about...](#)

I want to search by:

STRUCTURE

Draw, import MOL file, or type chemistry

OTHER PARAMETERS

Chemical name, CAS, species, sex, Lhasa TD₅₀, tumour

Type/paste InChI, SMILES or MOL

ENTER

IMPORT MOL FILE

H C N O S P F Cl Br I

[Clear all](#)

SEARCH



The Lhasa Carcinogenicity Database is a searchable repository of 6529 long-term carcinogenicity studies covering a total of 1529 chemicals. The database builds upon the work done between 1980 and 2005 by Lois Swirsky Gold and her team. For further details, please see <https://toxnet.nlm.nih.gov/cpdb>.

Some compounds from Gold *et al.* dataset have been combined for the Lhasa Carcinogenicity Database as they represent the same chemical entity (for example at technical grade and commercial grade).

[More about...](#)

I want to search by:

STRUCTURE

Draw, import MOL file, or type chemistry

OTHER PARAMETERS

Chemical name, CAS, species, sex, Lhasa TD₅₀, tumour

Search on

Individual studies ▾

Type the chemical name or CAS number

Species

All species

Asian Musk Shrew

Bush Baby

Cynomolgus Monkey

Dog

Gerbil

Guinea Pig

Hamster

Mouse

Rabbit

Rat

Rhesus Monkey

Syrian Hamster

Tree Shrew

Sex

All

♂

Mixed

♀

Female

♂

Male

Lhasa TD₅₀ from

to

(mg/kg/day)

Tumour site

Tumour type

[Clear all](#)

SEARCH 🔍



Search by

STRUCTURE

OTHER PARAMETERS

[< Back to results](#)

Captan



Summary

Species	Lhasa TD ₅₀ (mg/kg/day)	Gold TD ₅₀ (mg/kg/day)	Result	Sex	Tumour sites	Notes
Mouse	1,010	2,110	POSITIVE	♀ Female	Small intestine	
				♂ Male	Small intestine	
Rat	1,410	2,080	POSITIVE	♀ Female	Uterus	-
				♂ Male	Kidney	

Chemical structure

CAS Number Chemistry unique identifier

133-06-2 133-06-2

Chemistry name

Captan

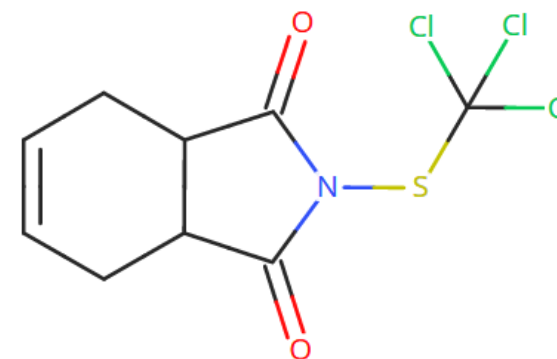
Synonym(s)

3a,4,7,7a-Tetrahydro-2-[(trichloromethyl)thio]-1H-isoindole-1,3(2H)-dione; N-Trichloromethylthio-4-cyclohexene-1,2-dicarboximide; N-Trichloromethylthiotetrahydrophthalimide; Orthocide; Trichloromethylthio-1,2,5,6-tetrahydrophthalimide

Molecular weight Molecular formula

300.59 C₉H₈Cl₃NO₂S

SMILES

C1C=CCC2C1C(N(C2=O)SC(Cl)(Cl)Cl)=O

Search by

STRUCTURE

OTHER PARAMETERS

Study details and citations

expand all | collapse all

Species	Sex	Strain	Route	Exposure time	Experiment time															
Mouse	Male	CD1	Diet	113 week(s)	113 week(s)	▼														
Tumour Site Small intestine	Tumour Type More than one tumour type		Probability ≤ 0.0005	POSITIVE	Lhasa TD ₅₀ 956	Gold TD ₅₀ 2690														
<table border="1"> <thead> <tr> <th>Unit</th> <th>mg/kg/day</th> <th>mg/kg/day</th> <th>mg/kg/day</th> <th>mg/kg/day</th> </tr> </thead> <tbody> <tr> <td>Dose</td> <td>0</td> <td>703</td> <td>1180</td> <td>1890</td> </tr> <tr> <td>Incidence</td> <td>3/80</td> <td>19/80</td> <td>22/80</td> <td>39/80</td> </tr> </tbody> </table>	Unit	mg/kg/day	mg/kg/day	mg/kg/day	mg/kg/day	Dose	0	703	1180	1890	Incidence	3/80	19/80	22/80	39/80					
Unit	mg/kg/day	mg/kg/day	mg/kg/day	mg/kg/day																
Dose	0	703	1180	1890																
Incidence	3/80	19/80	22/80	39/80																
Tumour Site Small intestine	Tumour Type Carcinoma		Probability ≤ 0.0005	POSITIVE	Lhasa TD ₅₀ 1450	Gold TD ₅₀ 3500														
Tumour Site Small intestine	Tumour Type Adenoma		Probability ≤ 0.002	POSITIVE	Lhasa TD ₅₀ -	Gold TD ₅₀ 8280														
Literature reference(s)						>														
Notes (exposure, histopathology, mortality)						>														
Species	Sex	Strain	Route	Exposure time	Experiment time															
Mouse	Female	CD1	Diet	113 week(s)	113 week(s)	▼														
Tumour Site Small intestine	Tumour Type Adenoma		Probability ≤ 0.035	POSITIVE	Lhasa TD ₅₀ -	Gold TD ₅₀ 13300														
Tumour Site Small intestine	Tumour Type Carcinoma		Probability ≤ 0.0005	POSITIVE	Lhasa TD ₅₀ 1320	Gold TD ₅₀ 2110														
Tumour Site Small intestine	Tumour Type More than one tumour type		Probability ≤ 0.0005	POSITIVE	Lhasa TD ₅₀ 1600	Gold TD ₅₀ 1710														

Search by

STRUCTURE


OTHER PARAMETERS

[< Back to results](#)

Capsaicin



Summary

Species	Lhasa TD ₅₀ (mg/kg/day)	Gold TD ₅₀ (mg/kg/day)	Result	Sex	Tumour sites	Notes
Mouse	-	167	POSITIVE	♀ Female	Caecum	
			POSITIVE	♂ Male	Caecum	

Chemical structure

CAS Number Chemistry unique identifier
404-86-4 404-86-4

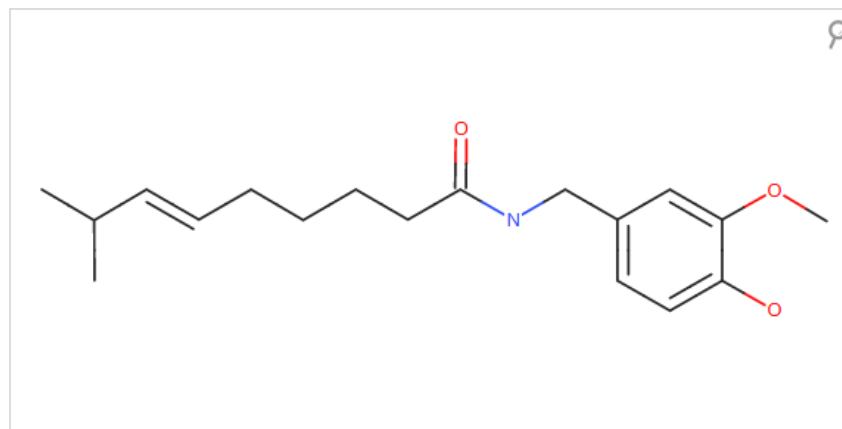
Chemical name
Capsaicin

Synonym(s)
(6E)-N-[(4-Hydroxy-3-methoxyphenyl)methyl]-8-methyl-6-nonenamide; (E)-8-Methyl-N-vanillyl-6-nonenamide; Isodecenoic acid vanillylamide; N-(4-Hydroxy-3-methoxybenzyl)-8-methylnon-trans-6-enamide; trans-8-Methyl-N-vanillyl-6-nonenamide

Molecular weight Molecular formula
305.41 C₁₈H₂₇NO₃

SMILES
C=1(C(=CC=C(=1)CNC(=O)CCCC=CC(C)C)O)OC

InChI



Search by

STRUCTURE

OTHER PARAMETERS

Study details and citations

expand all | collapse all


Species Mouse	Sex Female	Strain Swiss	Route Diet	Exposure time 152 week(s)	Experiment time 152 week(s)		▼						
Tumour Site Caecum	Tumour Type Adenoma- polypoid		Probability ≤ 0.148	POSITIVE	Lhasa TD ₅₀ -	Gold TD ₅₀ 167	▼						
<table border="1"> <thead> <tr> <th>Dose (mg/kg/day)</th> <th>0</th> <th>40.8</th> </tr> </thead> <tbody> <tr> <td>Incidence</td> <td>4/50</td> <td>11/50</td> </tr> </tbody> </table>								Dose (mg/kg/day)	0	40.8	Incidence	4/50	11/50
Dose (mg/kg/day)	0	40.8											
Incidence	4/50	11/50											
Tumour Site Duodenum	Tumour Type Adenocarcinoma		Probability ≤ 0.253	NOT SPECIFIED	Lhasa TD ₅₀ -	Gold TD ₅₀ 2310	▼						
<table border="1"> <thead> <tr> <th>Dose (mg/kg/day)</th> <th>0</th> <th>40.8</th> </tr> </thead> <tbody> <tr> <td>Incidence</td> <td>0/50</td> <td>1/50</td> </tr> </tbody> </table>								Dose (mg/kg/day)	0	40.8	Incidence	0/50	1/50
Dose (mg/kg/day)	0	40.8											
Incidence	0/50	1/50											
Tumour Site Liver	Tumour Type Tumour- hepatocellular		Probability = 1	NOT SPECIFIED	Lhasa TD ₅₀ -	Gold TD ₅₀ -	>						
Tumour Site Lung	Tumour Type Multiple tumour types		Probability ≤ 0.93	NOT SPECIFIED	Lhasa TD ₅₀ -	Gold TD ₅₀ 1080	>						
Tumour Site Lung	Tumour Type Adenoma		Probability ≤ 0.631	NOT SPECIFIED	Lhasa TD ₅₀ -	Gold TD ₅₀ 225	>						
Literature reference(s)							>						
Species Mouse	Sex Male	Strain Swiss	Route Diet	Exposure time 124 week(s)	Experiment time 124 week(s)		▼						
Tumour Site Caecum	Tumour Type Adenoma- polypoid		Probability = 1	POSITIVE	Lhasa TD ₅₀ -	Gold TD ₅₀ -	>						

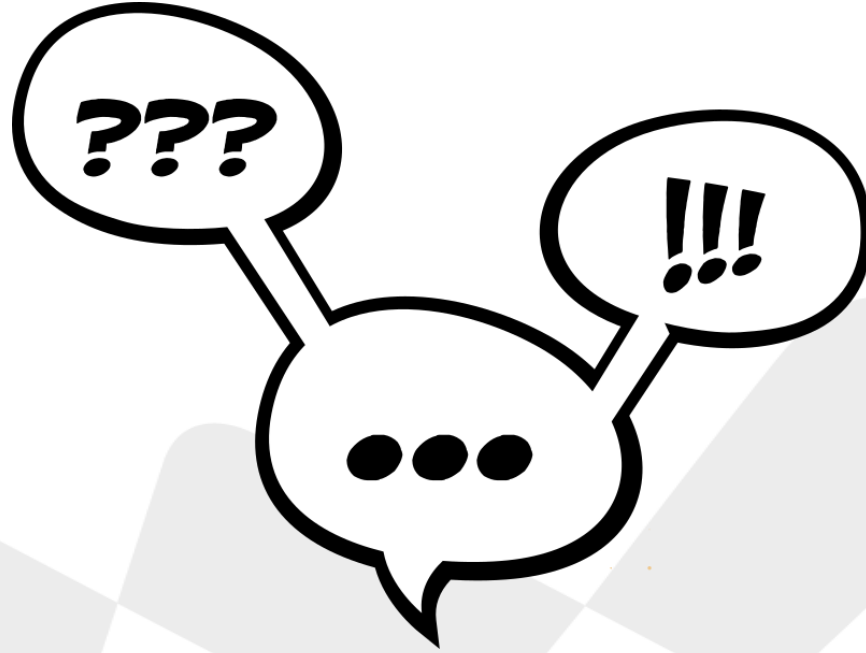
What does the future hold for Lhasa CarcDB?

- We want Lhasa CarcDB to be the new **Gold Standard** in carcinogenicity studies and calculated TD₅₀ values
- Short term goal
 - Increase usage and awareness of CarcDB
 - Create an interface where users can easily calculate their own TD₅₀ values using Lhasa's methodology



What does the future hold for Lhasa CarcDB?

- Adding data beyond the original Gold compounds
 - Additional data from the NTP dataset
 - Additional data from the literature
 - Investigating other measures of Carcinogenic potency to calculate and include alongside TD₅₀
 - Benchmark 10
 - Is there any feature *you* would like to see in CarcDB? We want your feedback.
- 



<https://carcdb.lhasalimited.org>



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

Company Registration Number 01765239. Registered in England and Wales. VAT Registration Number GB 396 8737 77.



ISO 9001: CERTIFIED