

# **Toxicological Data Resources: The Basis of Computational Modelling and Safety Assessment**

Andrea Richarz, Liverpool John Moores University

SEURAT-1 & ESTIV Joint Summer School 2014

9 June 2014, Egmond aan Zee

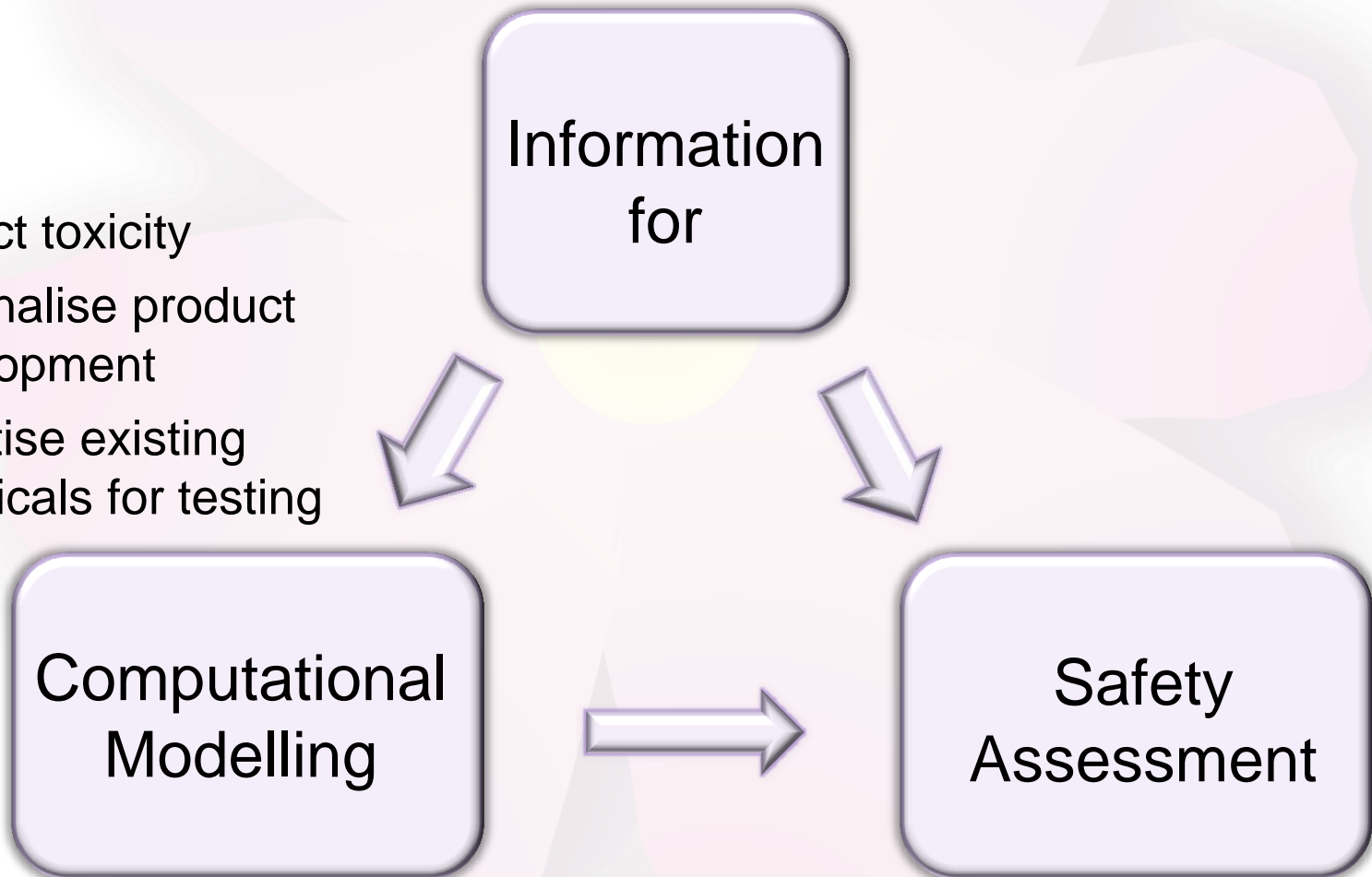


# Content

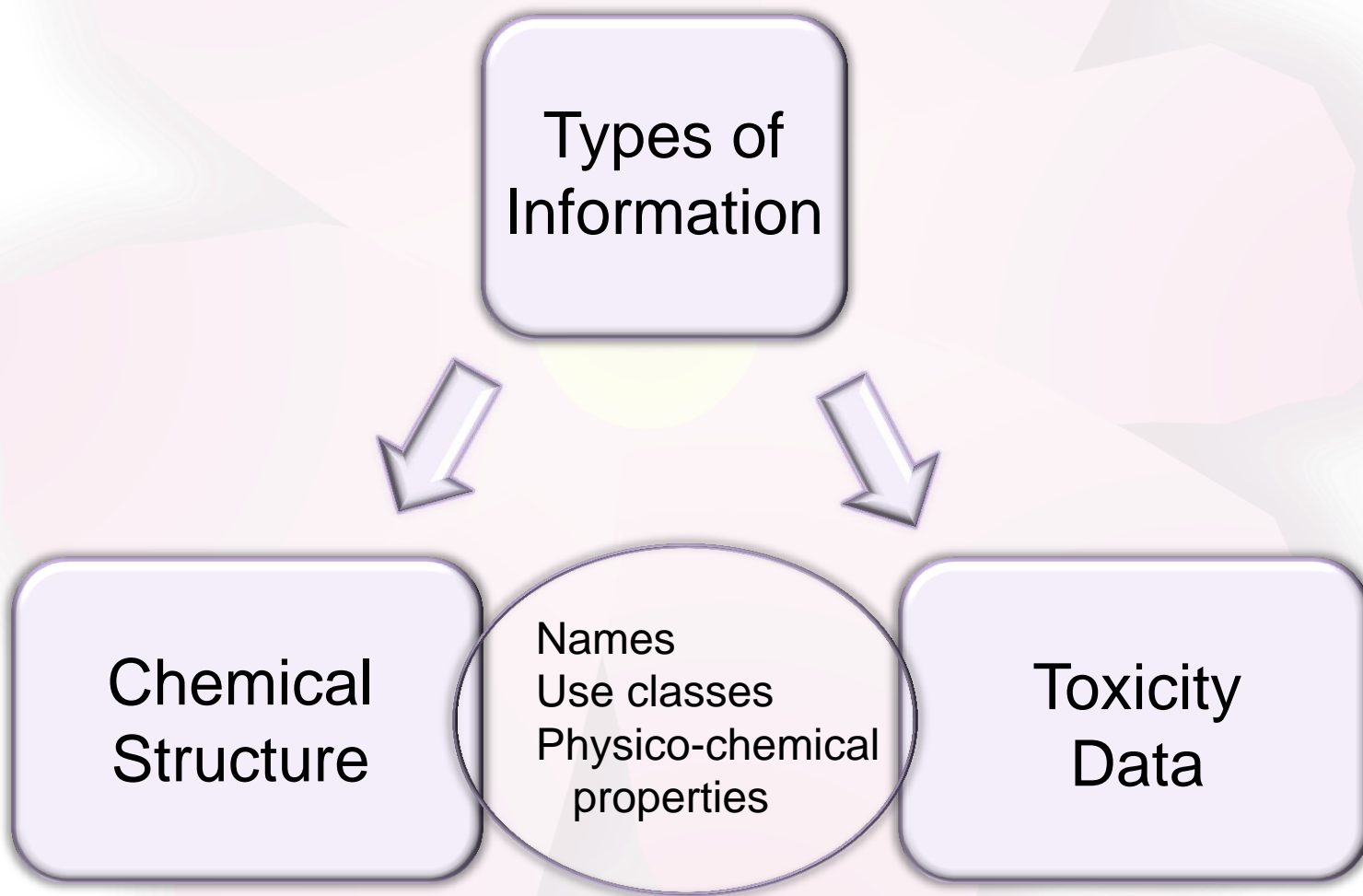
- Introduction to chemical structure representation
- Introduction to on-line data resources for chemical structure definition and toxicity data: Examples
- Issues of data quality and its importance for computational modelling
- Introduction to COSMOS DB: chemistry and toxicity data search
  - Find substances by name/structure/substructure
  - Find specific toxicity information

# Toxicological Data

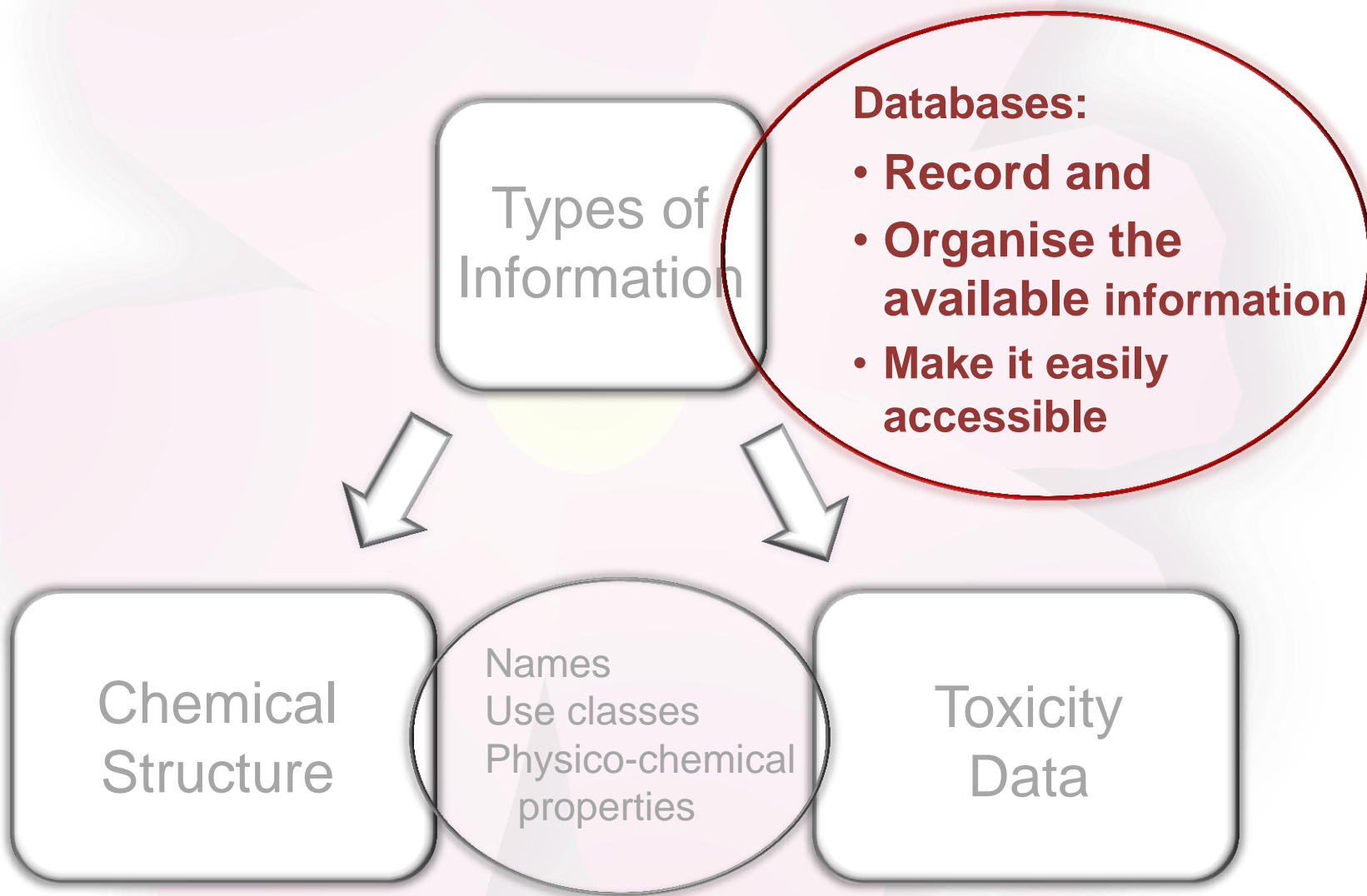
- Predict toxicity
- Rationalise product development
- Prioritise existing chemicals for testing



# Toxicological Data



# Toxicological Data

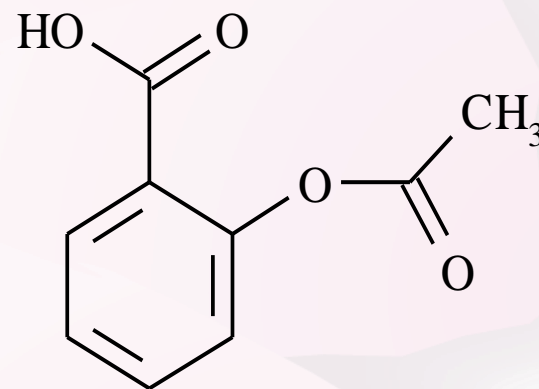


# Chemical Information

- Information regarding a defined chemical substance
  - Identity
  - Structure
  - Properties
  - Activities (beneficial and / or toxicity)
  - Uses
  - Cost, synthesis, manufacture
  - Others.....
- Can be stored and retrieved in a database

# Chemical Information for Aspirin

- White crystalline powder
- Acetylsalicylic acid
- $C_9H_8O_4$
- Hydrophilic (log P 1.4)
- Ionisable ( $pK_a$  3.5)
- Small (Mol Wt 180.2)
- Analgesic, anti-inflammatory
- GI toxicity
- Cheap, simple synthesis



# Chemical Information

- Information regarding a defined chemical substance
  - **Identity**
  - Structure
  - Properties
  - Activities (beneficial and / or toxicity)
  - Uses
  - Cost, synthesis, manufacture
  - Others.....
- Can be stored and retrieved in a database



# Chemical Name

- To link information to a chemical, its identity must be defined
- The structure of a chemical can be defined by its name
  - Common name
  - Trade Name
  - Chemical name
  - IUPAC name

# Chemical Name: Problems

- It may not be possible to obtain the structure from the name
- Not all names are specific to a single chemical structure (e.g. xylene, dichlorobenzene)
- Language differences
- Defining the chemical structure
  - Isomers
  - Salts
  - Mixtures (e.g. surfactants)

# Names for Aspirin (from ChemSpider)

2-(Acetyloxy)benzoic acid		acide acetylsalicylique	Easprin	Rhonal
2-(Acetyloxy)benzolcarbonsäure [German]		acido acetilsalicilico [Italian]	Easprin (TN)	ROBAXISAL
2-Acetoxybenzenecarboxylic acid	Salicylic acid, acetyl-	Acido O-acetil-benzoico [Italian]	Ecolen	rodine
2-Acetoxybenzoesäure [German]	Salospir	acidum acetylsalicylicum	Ecotrin	Ronal
2-Acetoxybenzoic acid [IUPAC Name]	Tasprin	Acimetten	EmpirinEndosprin	Salacatin
2-acetyloxybenzoic acid	Toldex	Acisal	Endydol	Salcetogen
50-78-2 <a href="#">[RN]</a>	Triaminicin	Acylpyrin	Entericin	Saletin
A.S.A.	阿司匹林	Adiro	Enterophen	Salicylic acid, acetate
Acesan	200-064-1 [EINECS]	AIN	Enterosarein	Solfrin
Acetard	2-Acetoxy-benzoic acid	Aloxiprium	Enterosarine	Solprin
Acetoxybenzoic acid	2-Carboxyphenyl acetate [Spanish]	Anacin	Entrophen	Solprin acid
acetyl salicylic acid	4-10-00-00138 [Beilstein]	ASA	EQUAGESIC	Solpyron
Acetyl-SAL	779271 [Beilstein]	Asagran	Extren	Solupsan
ACETYLSALICYLIC ACID	Acenterine	Ascriptin	Globentyl	Spira-Dine
Acetyonyl	Acesal	Aspec	Globoid	Supac
Acetylsalicylic acid	Aceticyl	Aspergum	Helicon	Synalgos
Acid, Acetylsalicylic	Acetilsalicilico	Aspirdrops	Idragin	Temperal
acide 2-(acétyloxy)benzoïque [French]	Acetilum acidulatum	Aspirin [BAN:JAN]	Istopirin	Triple-sal
Acide 2-acétoxybenzoïque [French]	Acetisal	<a href="#">Aspirine</a>	Kapsazal	Vanquish
Asatard	acetol	Aspro	Levius	Xaxa
Aspirin <a href="#">[Wiki]</a>	Acetophen	Aspro Clear	Magnecyl	Yasta
Aspirin (JP15/USP)	Acetosal	Asteric	MEASURIN	Zorprin
Aspirin (VAN)	Acetosalin	AXOTAL	Medisyl	
Aspropharm	ACETYL SALICYLIC ACID	Azetylsalizylsaeure	Micristin	
Benzoic acid, 2-(acetyloxy)-	Acetylin	Benaspir	Novid	
Cardioaspirina	Acetylsal	Benzoic acid, 2- (acetyloxy)-	o-acetoxybenzoic acid	
Doril	acetylsalicylate	Bialpirina	o-carboxyphenyl acetate	
ECM	acetylsalicylicacid	Bialpirinia	PERCODAN	
Kyselina 2-acetoxybenzoova [Czech]	Acetylsalicylsaeure	Bi-prin	Persistin	
Kyselina acetylsalicylova [Czech]	Acetylsalicylsaure [German]	Bufferin	Pharmacin	
Melhoral	acetyl-salicylsyra	Caprin	Pirseal	
Miniasal	acetylsalicylzuur	Cemirit	Polopirin	
o-(Acetyloxy)benzoic Acid	Acetylsaliylic acid	Claradin	Polopiryna	
o-acetylsalicylic acid	Acetylsalicylic acid	Clariprin	Pravigard PAC	
Rhodine NC RP	Acetysal	Colfarit	Premaspin	
	acide 2-(acetyloxy)benzoique	Contrheuma retard	Rheumin tabletten	
		Crystar	Rheumintabletten	
		D001241	Rhodine	
		Decaten		
		Delgesic		
		Dispril		
		Duramax		

# Chemical Identification Numbers

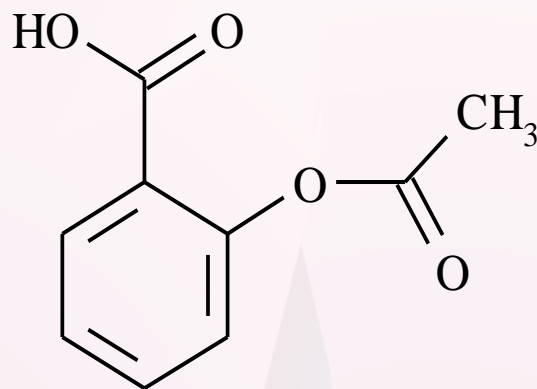
- Chemical Abstracts Service (CAS) Number
  - A CAS Registry Number is a numeric identifier that can contain up to 10 digits, divided by hyphens into three parts.
  - Aspirin: **50-78-2**
  - Over 60 million chemical structures (1957-present)
  - No chemical significance
  - [www.cas.org](http://www.cas.org)
- Problems:
  - Not unique
  - Copyrighted by American Chemical Society (ACS)

# Chemical Information

- Information regarding a defined chemical substance
  - Identity
  - **Structure**
  - Properties
  - Activities (beneficial and / or toxicity)
  - Uses
  - Cost, synthesis, manufacture
  - Others.....
- Can be stored and retrieved in a database

# Structure of Aspirin

- **Methods are required to translate a chemical structure into a format that can be entered into computer software and stored**
- **May be 2-D or 3-D (or even 4-D)**



# 2-D Chemical Structure Storage

- The simplest method to store a chemical structure
- Convert a chemical structure into a line of text
- Enables a computer to read a chemical structure
  - To store the structure
  - To perform calculations on the structure

# Simplified Molecular Input Line Entry Specification (SMILES)

- A **SMILES** string describes a chemical structure as line of text
- Developed in 1980s due to need to calculate log P
  - Freely available
  - Industry standard
- Easiest notation for a chemist to understand, visualise and write
- Many automated methods
- Can be converted to 3-D
- Easy to store for millions of compounds



# Using SMILES

- Can be written manually or using e.g. Java Molecular Editor (JME)
  - <http://www.molinspiration.com/cgi-bin/properties#>
  - Draw molecule, click 😊 to open a new window with a **SMILES** string
- Tutorials and further information
  - [http://www.epa.gov/med/Prods\\_Pubs/smiles.htm](http://www.epa.gov/med/Prods_Pubs/smiles.htm)
  - [http://www.daylight.com/dayhtml\\_tutorials/languages/smiles/index.html](http://www.daylight.com/dayhtml_tutorials/languages/smiles/index.html)

# IUPAC International Chemical Identifier (InChI™)

- Non-proprietary identifier for chemical substances
- A unique code which includes 2-D and 3-D information
- Difficult to write by hand, but can be created by free software (e.g. ACDLabs ChemSketch)
- More information:
  - <http://www.iupac.org/inchi/>
  - [www.inchi.info](http://www.inchi.info)

# InChiKey

- InChi codes are too long to search e.g. with Google
- A fixed-length (25-character) condensed digital representation of the InChi code is known as the InChiKey
- Not comprehensible

# 2-D Representation of the Structure of Aspirin

- **SMILES**

CC(=O)Oc1ccccc1C(=O)O

- **InChi**

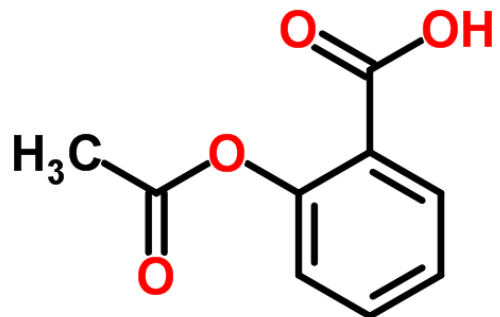
1S/C9H8O4/c1-6(10)13-8-5-3-2-4-7(8)9(11)12/h2-5H,1H3,(H,11,12)

- **InChiKey**

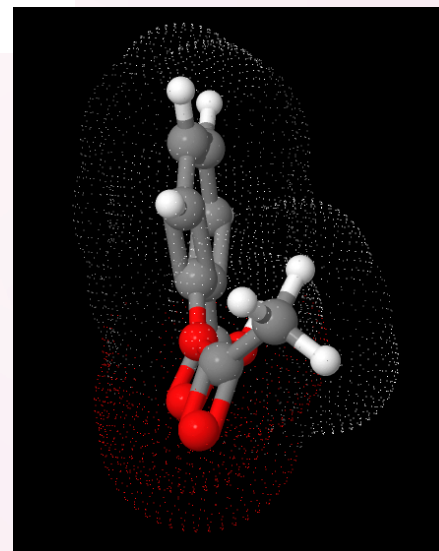
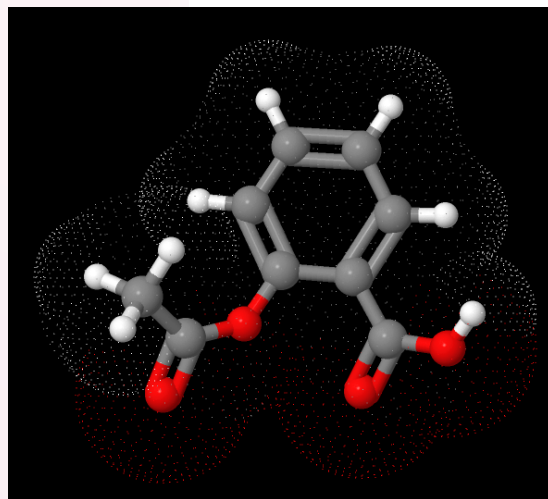
BSYNRYMUTXBXSQ-UHFFFAOYSA-N

# 2-D vs 3-D Chemical Structure Representation and Storage

2-D



3-D



Structures from ChemSpider

# 2-D vs 3-D Chemical Structure Storage

- 3-D representation requires different and much more complex method to store
  - Many formats e.g. .mol, .sdf, .pdb

# 2-D vs 3-D Chemical S

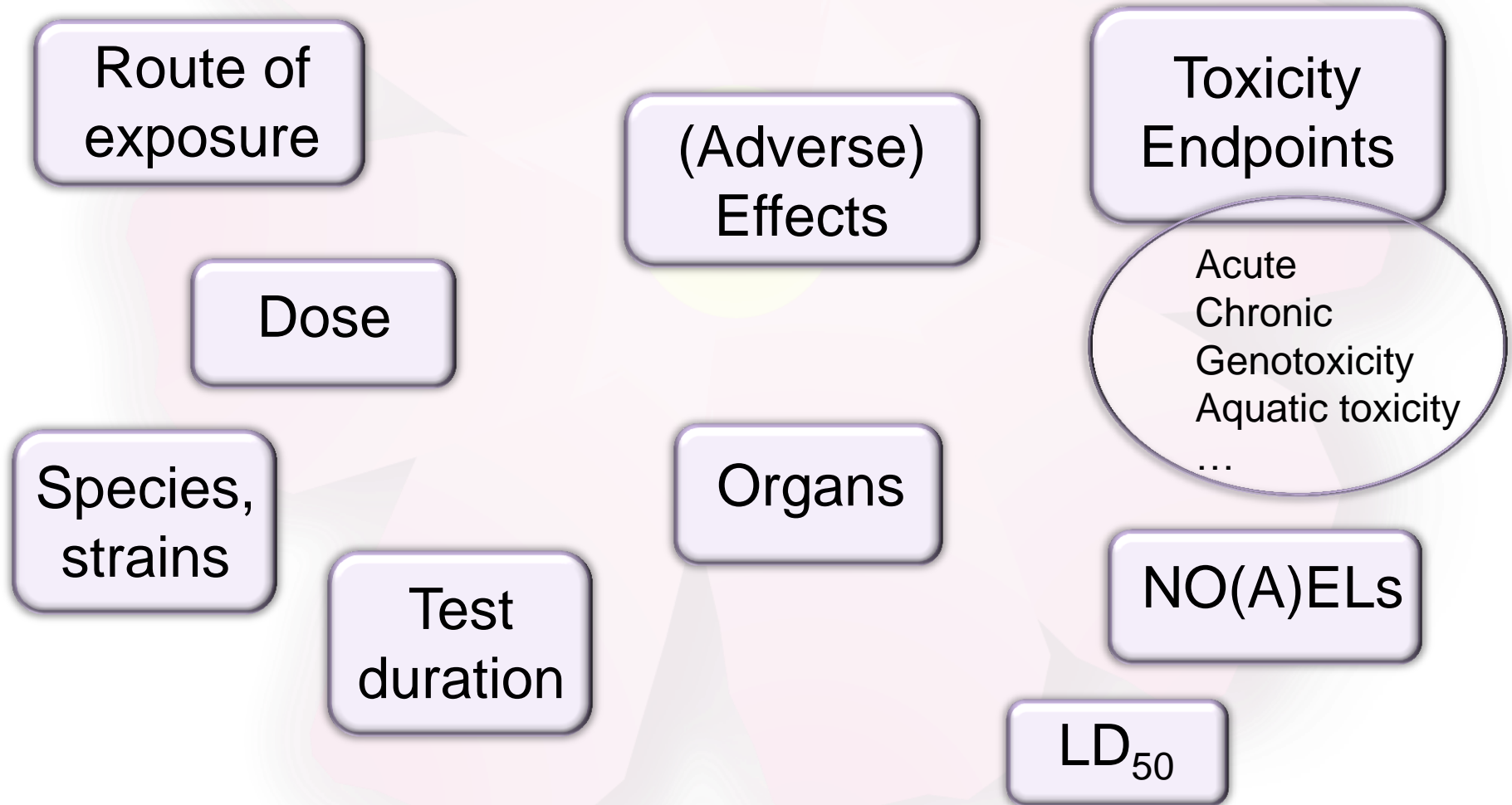
- 3-D representation require much more complex meth  
– Many formats e.g. .mol, .sd

**.mol file for aspirin**

```
http://www.chemspider.com/FilesHandler.ashx?type=str&3d=yes&id=2157
Jmol version 12.1.4_dev 2010-08-06 18:30
EXTRACT: ({0:20})
21 21 0 0 0
-0.913777 -1.359255 -0.662877 O
3.3608096 -1.604774 -0.048877 O
1.3941175 -2.618084 -0.330118 O
-2.755977 -2.455504 -0.102624 O
-0.197698 -0.178564 -0.529071 C
1.2887748 -0.257025 -0.332200 C
-0.768034 1.0369505 -0.684922 C
2.0094273 0.8848193 -0.280085 C
0.0428267 2.2685506 -0.603704 C
1.3575926 2.197486 -0.411373 C
2.000719 -1.560574 -0.229435 C
-2.056418 -1.486647 0.1333034 C
-2.565141 -0.590366 1.2421317 C
-1.826942 1.131077 -0.874311 H
3.0820649 0.8849562 -0.155539 H
-0.442427 3.2342796 -0.710428 H
1.9541266 3.1017172 -0.361621 H
-2.699147 -1.183326 2.1713195 H
-3.558291 -0.188477 0.9457143 H
-1.897960 0.2561982 1.4778912 H
3.5182405 -1.233845 0.85909 H
1 5 1
1 12 1
2 11 1
2 21 1
3 11 2
4 12 2
5 6 1
5 7 2
6 8 2
6 11 1
7 9 1
7 14 1
8 10 1
8 15 1
9 10 2
9 16 1
10 17 1
12 13 1
13 18 1
13 19 1
13 20 1
```

# Toxicity Information in Databases

Mostly: Existing information from *in vivo* tests





# ChemSpider:

[www.chemspider.com](http://www.chemspider.com)

- **Over 30 million compounds** from over 470 data sources
- Many high quality structures – **aim to improve the quality of public chemistry data sources**
- Easy and powerful search options
- Structure and substructure search
- Properties & data
- Funded through RSC and commercial links

# http://www.chemspider.com



[RSC](#) [SyntheticPages](#) [Login](#) [Register](#)

**RSC** | Advancing the  
Chemical Sciences

[About](#) | [More Searches](#) | [Web APIs](#) | [Help](#)

eg. Pyridine

**Search**

**Simple search** | [Structure search](#) | [Advanced search](#)

dimethicone

Systematic names  
1,2-dihydroxybenzene

Synonyms  
AIBN

Trade names  
Aspirin

Registry numbers  
7732-18-5

SMILES  
O=C(OCC)C

**Search**

Type Compound Name Here and  
then Click on "Search"

## What is ChemSpider?

ChemSpider is a free chemical structure database providing fast text and structure search access to over 26 million structures from hundreds of data sources. Watch [our introduction video](#).

## Search by chemical names

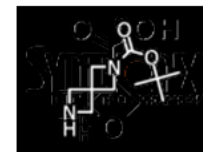
- Systematic names
- Synonyms
- Trade names
- Database identifiers

## Search by chemical structure

- Create structure-based queries
- Draw structures in the web page
- Use structure files from your computer

## Find important data

- Literature references
- Physical properties
- Interactive spectra
- Chemical suppliers



# ChemSpider for Aspirin

- Names and identifiers
  - Validated and non-validated, 2-D / 3-D structure
- Properties
  - Physical and chemical: experimental / predicted
- Spectra
- Literature
- Vendors
- Data sources
- Wikis
- Patents
- Pharmacological links
- Many more ....

# ToxNet:

[toxnet.nlm.nih.gov](http://toxnet.nlm.nih.gov)

- TOXNET (TOXicology Data NETwork) is a **cluster of databases** covering toxicology, hazardous chemicals, environmental health and related areas.

- Hazardous Substances Data Bank
- Integrated Risk Information System
- International Toxicity Estimates for Risk
- Chemical Carcinogenesis Research Information System
- GENE-TOX (Genetic Toxicology)
- Tox Town®
- Household Products Database
- Haz-Map®
- TOXMAP®
- LactMed (Drugs and Lactation)
- Carcinogenic Potency Database
- Comparative Toxicogenomics Database
- TOXLINE
- Development and Reproductive Toxicology/Environmental Teratology Information Center
- Toxics Release Inventory (TRI)
- ChemIDplus

# ToxNet for Aspirin



United States  
National Library  
of Medicine

## TOXNET Toxicology Data Network



[SIS Home](#) | [About Us](#) | [Site Map & Search](#) | [Contact Us](#)

[Env. Health & Toxicology](#) [TOXNET](#)

**TOXNET** - Databases on toxicology, hazardous chemicals, environmental health, and toxic releases.

### Select Database

- [ChemIDplus](#) [?](#)
- [HSDB](#) [?](#)
- [TOXLINE](#) [?](#)
- [CCRIS](#) [?](#)
- [DART](#) [?](#)
- [GENETOX](#) [?](#)
- [IRIS](#) [?](#)
- [ITER](#) [?](#)
- [LactMed](#) [?](#)
- [Multi-Database](#) [?](#)
- [TRI](#) [?](#)
- [Haz-Map](#) [?](#)
- [Household Products](#) [?](#)
- [TOXMAP](#) [?](#)

### Additional Resource

- [CPDB](#) [?](#)
- [CTD](#) [?](#)

### Env. Health & Toxicology



[VISIT SITE](#)

Portal to  
environmental  
health and  
toxicology  
resources.

### Support Pages

[Help](#)

### Search All Databases

(e.g. asthma air pollution, ibuprofen fever, vinyl chloride)

### References from Biomedical Literature

TOXLINE	Toxicology Literature Online	22351
DART	Developmental Toxicology Literature	886

### Chemical, Toxicological, and Environmental Health Data

ChemIDplus	Chemical Identification/Dictionary	1
HSDB	Hazardous Substances Data Bank	219
CCRIS	Chemical Carcinogenesis Information	1
CPDB	Carcinogenic Potency Database	2
GENETOX	Genetic Toxicology Data	1
CTD	Comparative Toxicogenomics Database	586
IRIS	Integrated Risk Information	0
ITER	International Toxicity Estimates for Risk	0
LactMed	Drugs and Lactation Database	9
TRI	Toxics Release Inventory	0
TOXMAP	Environmental Health e-Maps	0
Haz-Map	Occupational Exposure/Toxicology	Show me
Household Products	Health & Safety Information on Household Products	0


# ChemIDplus:

<http://chem.sis.nlm.nih.gov/chemidplus>

- Database of over 370,000 chemicals
- US National Institutes of Health
- Searched by:
  - Chemical Name, CAS Registry Number, Molecular Formula, Classification Code, Locator Code, and Structure or Substructure
- Include search and display by Toxicity indicators such as Median Lethal Dose (LD<sub>50</sub>), by physical/chemical properties such as log P, and by molecular weight
- **Very useful to find / verify chemical structures**

# ChemIDplus

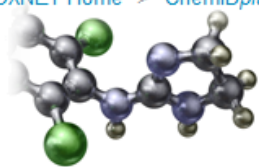
<http://chem.sis.nlm.nih.gov/chemidplus>

 **U.S. National Library of Medicine**

**TOXNET** TOXICOLOGY DATA NETWORK

[Help](#) | [FAQs](#) | [TOXNET Fact Sheet](#) | [Training Manual & Schedule](#)

[TOXNET Home](#) > [ChemIDplus Lite](#) • [Browse ChemIDplus](#) • [ChemIDplus Advanced](#)

**ChemIDplus**  
A TOXNET DATABASE

[Search](#) [Clear](#) [History](#) [Help](#)

Display  results

**Substance Identification** ⓘ  
Name/Synonym  Equals   
  
Data is available for 402,507 records.

**Toxicity** ⓘ  
Test: (any)  between  (mg/kg or ppm)  
Species: (any)   
Route: (any)   
Effect: (any)   
Toxicity data is available for 139,354 records.

**Physical Properties** ⓘ  
Melting Point   
between   
Either  Measurement Type  
Physical property data is available for 25,461 records  
and was provided by [Syracuse Research Corporation](#).

**Locator Codes** ⓘ

**Structure** ⓘ  
[Draw](#)  
Powered by [ChemAxon Marvin](#)  
Use:   
[Import MOL](#)  
**Structure Search Options** ⓘ  
☐ Substructure Search  
☒ Similarity Search  %  
☐ Exact (parent only)  
☐ Flex (parent, salts, mixture)  
☐ Flexplus (parent, all variations)  
Structure data is available for 313,456 records.

**Molecular Weight** ⓘ  
between   
Molecular weight data is available  
for 313,456 records.



# ChemIDplus for Aspirin


Organism	Test Type	Route	Reported Dose (Normalized Dose)	Effect	Source
child	LDLo	oral	104mg/kg (104mg/kg)	LUNGS, THORAX, OR RESPIRATION: ACUTE PULMONARY EDEMA GASTROINTESTINAL: NAUSEA OR VOMITING BLOOD: HEMORRHAGE	Lancet. Vol. 2, Pg. 809, 1952. <a href="#">Link to PubMed</a>
child	TDLo	oral	10mg/kg/1D-I (10mg/kg)	LUNGS, THORAX, OR RESPIRATION: ACUTE PULMONARY EDEMA KIDNEY, URETER, AND BLADDER: "CHANGES IN TUBULES (INCLUDING ACUTE RENAL FAILURE, ACUTE TUBULAR NECROSIS)" KIDNEY, URETER, AND BLADDER: URINE VOLUME DECREASED	Clinical Toxicology. Vol. 18, Pg. 247, 1981. <a href="#">Link to PubMed</a>
child	TDLo	oral	39mg/kg/13D-I (39mg/kg)	LIVER: "HEPATITIS (HEPATOCELLULAR NECROSIS), DIFFUSE"	American Journal of Diseases of Children. Vol. 139, Pg. 453, 1985. <a href="#">Link to PubMed</a>
dog	LD50	intravenous	681mg/kg (681mg/kg)	BEHAVIORAL: ANALGESIA	Archives Internationales de Pharmacodynamie et de Therapie. Vol. 149, Pg. 571, 1964. <a href="#">Link to PubMed</a>
dog	LD50	oral	700mg/kg (700mg/kg)	BEHAVIORAL: ALTERED SLEEP TIME (INCLUDING CHANGE IN RIGHTING REFLEX) LUNGS, THORAX, OR RESPIRATION: RESPIRATORY DEPRESSION	Arzneimittel-Forschung. Drug Research. Vol. 21, Pg. 719, 1971. <a href="#">Link to PubMed</a>
guinea pig	LD50	oral	1075mg/kg (1075mg/kg)	BEHAVIORAL: ALTERED SLEEP TIME (INCLUDING CHANGE IN RIGHTING REFLEX) BEHAVIORAL: TREMOR BEHAVIORAL: SOMNOLENCE (GENERAL DEPRESSED ACTIVITY)	Journal of the American Pharmaceutical Association, Scientific Edition. Vol. 47, Pg. 479, 1958.
hamster	LD50	oral	3500mg/kg (3500mg/kg)		Archives of Toxicology, Supplement. Vol. 7, Pg. 365, 1984. <a href="#">Link to PubMed</a>
human	TDLo	oral	480mg/kg/7D-I (480mg/kg)	SENSE ORGANS AND SPECIAL SENSES: TINNITUS: EAR BEHAVIORAL: SOMNOLENCE (GENERAL DEPRESSED ACTIVITY) GASTROINTESTINAL: OTHER CHANGES	Arzneimittel-Forschung. Drug Research. Vol. 25, Pg. 281, 1975. <a href="#">Link to PubMed</a>
human	TDLo	oral	669mg/kg/11D (669mg/kg)	LIVER: LIVER FUNCTION TESTS IMPAIRED	American Journal of Hospital Pharmacy. Vol. 35, Pg. 330, 1978. <a href="#">Link to PubMed</a>
human	TDLo	oral	1050mg/kg/14D (1050mg/kg)	VASCULAR: OTHER CHANGES	Clinical Pharmacology and Therapeutics Vol. 67, Pg. 530, 2000. <a href="#">Link to PubMed</a>
human	TDLo	oral	2880mg/kg/8W (2880mg/kg)	SENSE ORGANS AND SPECIAL SENSES: TINNITUS: EAR GASTROINTESTINAL: NAUSEA OR VOMITING GASTROINTESTINAL: DECREASED MOTILITY OR CONSTIPATION	Arzneimittel-Forschung. Drug Research. Vol. 33, Pg. 631, 1983. <a href="#">Link to PubMed</a>



# ACToR

<http://actor.epa.gov/actor/faces/ACToRHome.jsp>

- Aggregated Computational Toxicology Resource
- **Meta-database** of US EPA databases (e.g. ToxRefDB, ToxCastDB, DSSTox); 500 000 chemicals; 1000 sources

 United States Environmental Protection Agency  
LEARN THE ISSUES | SCIENCE & TECHNOLOGY | LAWS & REGULATIONS | ABOUT EPA

ALL EPA THIS AREA Advanced Search  
 SEARCH

ACToR [Contact Us](#) [Share](#)

You are here: [EPA Home](#) » [Computational Toxicology Research](#) » [ACToR](#) » Home

ACToR ToxRefDB ToxCastDB ExpoCastDB DSSTox CSS Dashboards CPCat

[Home](#) | [Basic Info](#) | [Data Collections](#) | [Structure Search](#) | [Assays By Toxicity](#) | [Assays By Category](#) | [External Links](#) | [Download](#) | [Help](#)

ACToR (Aggregated Computational Toxicology Resource) is EPA's online warehouse of all publicly available chemical toxicity data and can be used to find all publicly available data about potential chemical risks to human health and the environment. ACToR aggregates data from over 1000 public sources on over 500,000 environmental chemicals searchable by chemical name, other identifiers and by chemical structure.

The data warehouse:

- Allows users to search and query data from other EPA chemical toxicity databases including:
  - ToxRefDB (30 years and \$2 billion worth of animal toxicity studies).
  - ToxCastDB (data from screening 1,000 chemicals in over 500 high-throughput assays).
  - ExpoCastDB (consolidate and link human exposure and exposure factor data for chemical prioritization).
  - DSSTox (provides high quality chemical structures and annotations).
- Includes chemical structure, physico-chemical values, in vitro assay data and in vivo toxicology data.
- Includes, but not limited to, high and medium production volume industrial chemicals, pesticides (active and inert ingredients), and potential ground and drinking water contaminants.

Chemical Name Parameters  
☒ Search on Chemical Names  
☐ Search on CAS Numbers

Match by  
☐ exact  
☒ any

Enter Chemical Name:

Search

Search Results

[Details](#) [Image](#) [CASRN](#) [Preferred Name](#) [Hazard](#) [Chronic Toxicity](#) [Carcinogenicity](#) [Genotoxicity](#) [Developmental Toxicity](#) [Reproductive Toxicity](#) [Food Safety](#) [Biomonitoring](#)

# OECD eChemPortal

<http://www.echemportal.org>

- **Meta-database**
- eChemPortal provides free public access to information on properties of chemicals:
  - Physical Chemical Properties
  - Ecotoxicity
  - Environmental Fate and Behaviour
  - Toxicity
- Simultaneous searching of reports and datasets
- Direct links to collections of chemical hazard and risk information prepared for **government chemical review programmes** at national, regional and international levels

# OECD eChemPortal

## http://www.echemportal.org

The Global Portal to Information on Chemical Substances



eChemPortal

eChemPortal



> Home

> Substance Search

> Property Search

> What's new?

> General Information

> Participating Databases

> Roles & Responsibilities

> Extension of the Portal

> Linking to eChemPortal

> Schedules of Assessments

> Structure Search

> GHS Classifications

> Other useful information

> FAQ

> Help

> Contact us

> Disclaimer

## Participating Databases

- Databases currently participating in eChemPortal
- Data sources which can be found through a search by Property
- Number of substance identity and endpoint records per participating source searchable through eChemPortal\*

### *Databases currently participating in eChemPortal:*

- **ACToR**  
U.S. EPA Aggregated Computational Toxicology Resource
- **AGRITOX**  
AGRITOX - Base de données sur les substances actives phytopharmaceutiques
- **APVMA-CR**  
The Australian Pesticides and Veterinary Medicines Authority (APVMA) database of completed chemical reviews
- **CCR**  
Canadian Categorization Results
- **CESAR**  
Canada's Existing Substances Assessment Repository
- **Combined Exposures**  
Collection of Case Studies on Risk Assessments of Combined Exposures to Multiple Chemicals
- **ECHA CHEM**  
European Chemicals Agency's Dissemination portal with information on chemical substances registered under REACH.
- **EnviChem**  
Data Bank of Environmental Properties of Chemicals
- **EPA HHBP**  
EPA Human Health Benchmarks for Pesticides
- **EPA OPPALB**  
EPA Office of Pesticide Programs' Aquatic Life Benchmarks
- **ESIS**  
European Chemical Substances Information System (ESIS)
- **GDL**  
Gefahrstoffdatenbank der Länder (Germany)
- **CHC 1**

k

# OECD eChemPortal

<http://www.echemportal.org>



Print

English ▼

## The Global Portal to Information on Chemical Substances



### eChemPortal ▼

- Home
- Substance Search
- Property Search
- What's new?
- General Information
- Participating Databases
- Roles & Responsibilities
- Extension of the Portal
- Linking to eChemPortal
- Schedules of Assessments
- Other useful information
- FAQ
- Help
- Contact Us
- Disclaimer

Chemical Substance  
Search

Chemical Property  
Data Search

Twenty  
Four data  
The list of data

Click on Box  
Type Compound and Click on  
"Search"

eChemPortal provides free public access to chemical information:

- Physical Chemical Properties
- Environmental Fate and Behaviour
- Ecotoxicity
- Toxicity

eChemPortal allows simultaneous searching of reports and datasets by chemical name and number and by chemical property. Direct links to collections of chemical hazard and risk information prepared for government chemical review programmes at national, regional and international levels are obtained. Classification results according to national/regional hazard classification schemes or to the Globally Harmonized System of Classification and Labelling of Chemicals (GHS) are provided when available. In addition, eChemPortal provides also exposure and use information on chemicals.

SPIN - Substances in Preparations In the Nordic countries Database has been added to eChemPortal

09 May 2012

Two additional databases:  
Combined exposure and Australian NICNAS  
Other available on eChemPortal

20 December 2011

### How to Use eChemPortal

Under Chemical Substance Search you can search for information by chemical number (recommended) or by

# CosIng

<http://ec.europa.eu/consumers/cosmetics/cosing>

- CosIng is the European Commission database with information on cosmetic substances and ingredients contained in the EC Cosmetics Regulation, Scientific Opinions on Cosmetic Substances of the Scientific Committees
- Notice:
  - Ingredients with an INCI name in CosIng are not necessarily actually used in cosmetic products nor approved for such use
  - For colorants, preservatives and UV filters, only those authorised in Annexes IV, VI, VII to Directive 76/768/EEC are listed in CosIng

# http://ec.europa.eu/consumers/cosmetics/cosing

## Use: Simple Search

into account all the characteristics of the product .

In addition to the above, the use of an ingredient in cosmetic products must be supported by the safety assessment of the product.

### About the search

**You can choose to search either in the Cosmetics Directive or in the Cosmetics Regulation. Please notice, that the annexes for the Regulation is not yet complete.**


You can search for the name of a substance (displayed in small letters) as it is referred to in the Cosmetics Regulation/Directive or for the name of an **INGREDIENT** (displayed in CAPITAL LETTERS), listed in the Inventory for labelling purposes, or for the name of a **FRAGRANCE** also in the Inventory.

Cosing allows also users to search for relevant CAS and EC numbers.

The current data in the database can be found under the default status as "active", whereas historical data have the status "not active".

Version

Cosmetics Directive ▼

Name  or  
CAS/EC #

dimethicone

Scope

All ▼

Status

Active ▼

Search

Type Compound Name Here and  
then Click on "Search"

or go to the:

- [Advanced search](#)
- [Proposed updates for the Inventory](#)
- Cosing's [Reference Data](#) (Regulations, Directives, Annexes, Functions, Abbreviations)

[Please keep us informed of any problems or requests](#)



# Databases of Repeated Dose Toxicity Data

- Fraunhofer ITEM RepDose  
<http://www.fraunhofer-repdose.de>
- ToxRefDB (US EPA)  
<http://www.epa.gov/ncct/toxrefdb>
- OECD QSAR Toolbox  
<http://www.qsartoolbox.org>
- NITE HESS RDT DB  
<http://www.safe.nite.go.jp/english/kasinn/qsar/hess-e.html>
- COSMOS DB  
<http://cosmosdb.cosmostox.eu>





# Other Resources

- Scientific Committee on Consumer Safety (SCCS) –  
Opinions  
[http://ec.europa.eu/health/scientific\\_committees/consumer\\_safety](http://ec.europa.eu/health/scientific_committees/consumer_safety)
- Priority-based Assessment of Food Additives (PAFA)  
database; US FDA  
<http://www.accessdata.fda.gov/scripts/fcn/fcnNavigation.cfm?rpt=iaListing>



# Look Up Ingredients in Different Databases

## Shampoo

- Aqua, sodium laureth sulfate, sodium lauryl sulfate, dimethicone, cocamide MEA, Zinc carbonate, glycol distearate, sodium chloride, zinc pyrithione, sodium xylenesulfonate, cetyl alcohol, parfum, guar hydroxypropyltrimonium chloride, magnesium sulfate, sodium benzoate, ammonium laureth sulfate, magnesium carbonate hydroxide, linalool, butylphenyl methylpropional, limonene, hydroxyisohexyl 3-cyclohexene carboxaldehyde, benzyl alcohol, hexyl cinnamal, citronellol, tocopheryl acetate, paraffinum liquidum, sodium polynaphthalenesulfonate, CI 19140, DMDM hydantoin, CI 15510, methylchloroisothiazolinone, disodium EDTA, tetrasodium EDTA, methylisothiazolinone

## Hand wash

- Aqua, sodium laureth sulfate, Cocamidopropyl betaine, sodium chloride, glycerin, polyquaternium-7, parfum, jasminum officinale, viola odorata flower extract, lactic acid, sodium benzoate, tetrasodium glutamate diacetate, propylene glycol, polysorbate 20, benzotriazolyl dodecyl p-cresol, benzyl sylicylate, alpha-isomethyl ionone, geraniol, hexyl cinnamal, butylphenyl methylpropional, linalool, hydroxyisohexyl 3-cyclohexene carboxaldehyde, CI 14700, CI 60730

ChemSpider: <http://www.chemspider.com/>

eChemPortal: <http://www.echemportal.org/>

CosIng: <http://ec.europa.eu/consumers/cosmetics/cosing/>

# Using Existing Data

## ***What do you need data for?***

Risk assessment?  
Screening?  
Modelling?

## ***What data are relevant?***

Which endpoint?  
Which species?  
Which time points?  
Which vehicle(s)?  
Single or repeat dose?

**Factors to  
Consider**

***Are the data correct  
and accurately  
curated?***

***How do you assess  
the quality of the  
data?***



***Are the data “fit for purpose”?***

# Caveats in Using Existing Data

*A model is only as good as the data  
on which it is built*

- Reliability depends on source:
  - In-house data – usually reliable and traceable
  - Literature – variable
  - Data compilations – increased transcription error
  - Databases - *up to 10%* of entries incorrect
- Biological and experimental variability increase errors
- Data quality assessment required<sup>#1</sup>
  - e.g. using Klimisch criteria<sup>#2</sup> to ensure suitable quality
  - **ToxRTool** (<http://ecvam.jrc.it/>)

# It is Essential to Have the Correct Chemical Structure

- Many databases contain errors in chemical structure
- Problems with...
  - Names
  - Tautomers
  - Stereoisomers
  - Salts
  - Mixtures
  - Conformational shape
  - etc
- INCHI Keys attempt to resolve problems
  - <http://www.inchi-trust.org>

# Finding Chemical Structures and Toxicity Data: Conclusions

**Many data are available to inform modelling/safety assessment**

Important aspects:

- What is the correct chemical structure?
  - Other issues e.g. salts, mixture
- What toxicity data are available? What purpose?
- Are the data correct?

**--> Quality and relevance need to be ascertained**

# Register for a Free COSMOS DB Account

- Free registration is available at <http://cosmosdb.cosmostox.eu/accounts/register> , setting up an individual login and password
- The password can be changed when logged in



# The Need for An Inventory and Toxicity Data Storage: **COSMOS DB**

- Summary of Key Motivations
  - Lack of a single inventory of cosmetics ingredients, associated with high quality chemical structures
  - Platform for extending the Threshold of Toxicological Concern (TTC) approach for cosmetics
  - Create comprehensive source for quality controlled data, especially for repeated dose toxicity, focussing on cosmetics-related substances
  - Knowledge creation by profiling the effects of repeated dose studies



# Public Release of COSMOS DB v1.0

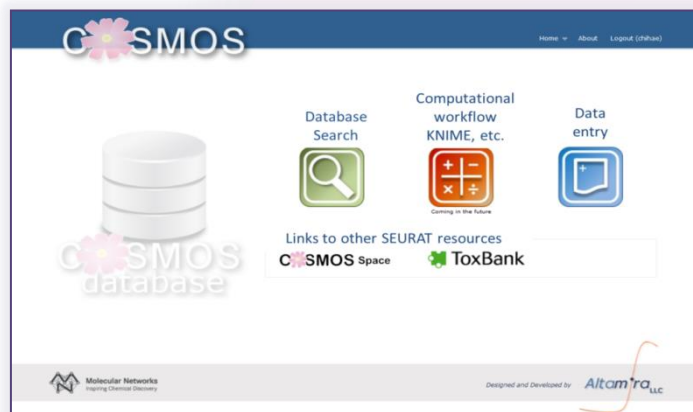
<http://cosmosdb.cosmostox.eu/>



- Search database
- Intuitively browse complex data



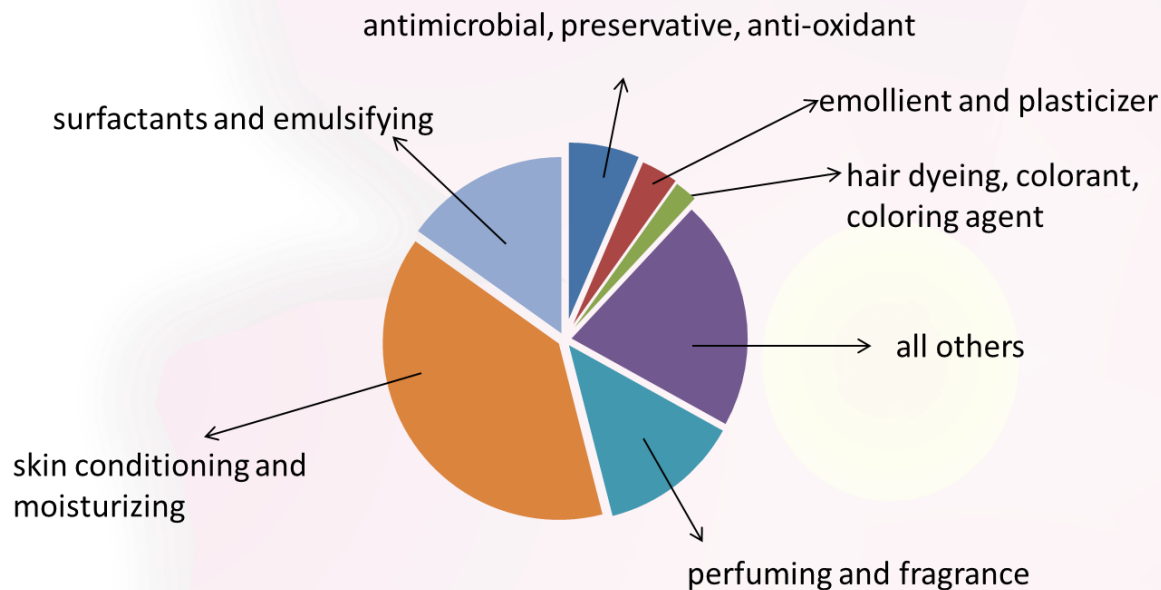
# COSMOS Database



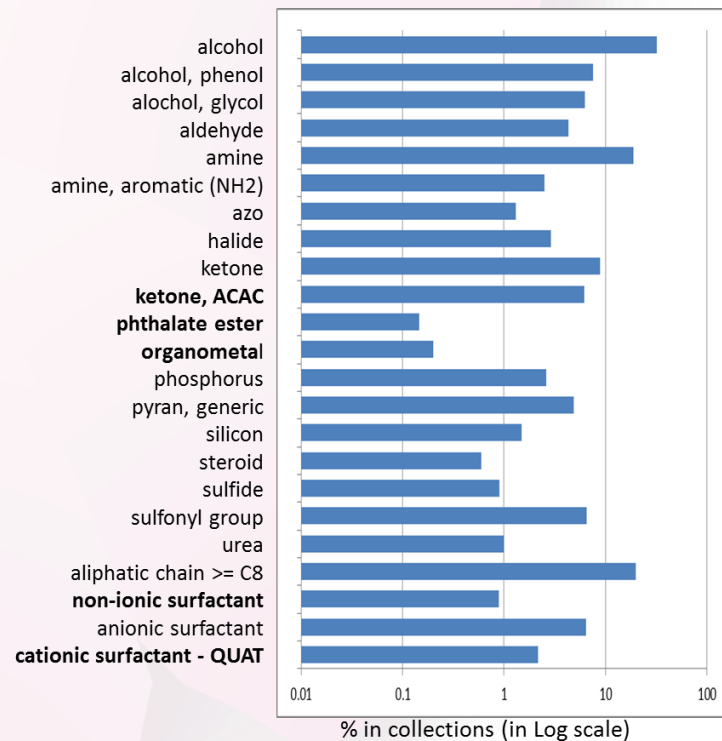
- Completely open-access technology
- High-quality toxicity data (quality control, structure curation)
- User-friendly query builder for retrieval of data by combining various criteria (toxicity study endpoint, animal species, strain, sex and/or route of exposure utilised in studies)



# COSMOS Cosmetics Inventory: Over 19,000 Substances – Use and Chemical Classes



Cosmetics Inventory associated with 66 unique use functions.  
Top 6 categories shown.

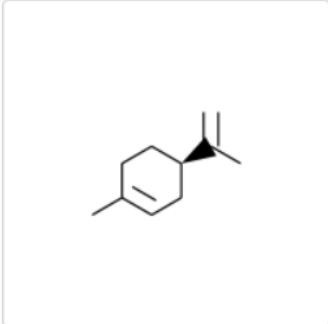


Chemical classes

# Content of COSMOS DB v1.0

- 81,602 COSMOS IDs
- 44,765 unique structures
- Cosmetics Inventory
  - INCI names
  - Chemical functions
- 12,538 toxicity studies for 1,660 compounds across 27 endpoints
- oRepeatTox DB toxicity data (230 cosmetics related chemicals and 340 studies)
- US FDA PAFA post-market and toxicity
- ...further data to come

Structure #1



Stereochemistry	unassigned
Double Bond Geometry	unassigned
Structure Source	SF/Registry
Structure Quality	High
Structure Representation	actual

+ IDs and Names

- Compound Annotations


Formula	
Material Type	ORGANIC
Composition Type	DEFINED FORMULA
Comments	
Use Types	FLAVOR
Use Functions	<a href="#">MASKING [COSING]</a> <a href="#">PERFUMING [COSING]</a> FLAVORING AGENT OR ADJUVANT [US FDA CFSAN PAFA]
Inventories	COSING : <a href="#">link 1</a> PCPC

+ PAFA Chemical Information

+ Toxicity Data

# Searching the Database

- Searching for a substance/list of substances:
  - Name
  - CAS number
  - Identifiers
  - Structure:
    - Sketch structure
    - SMILES strings
- Exact/Partial/Similar



**COSMOS** Database Search

Query Definitions: Chemistry

Names Identifiers CAS Registry Number **Structure: Sketch molecule** Structure: SMILES

Edit molecule Clear



SMILES: C1CCCC1

Options

☐ exact ☒ partial ☐ similar

Molecular Networks  
Inspiring Chemical Discovery

# Searching the Database

- Toxicity study parameter search

The screenshot displays the COSMOS Database Search interface. A modal dialog box titled "Add Tox Query" is open, allowing users to define search parameters for toxicity studies. The background interface includes a header with the COSMOS logo and navigation links (Home, About, alexandra.ms), a sidebar for "Query Definitions: Chemistry" with tabs for "Names" and "Identifiers", and a "Search" button at the bottom right.

**Add Tox Query**

**Endpoint**  
Cytotoxicity (731)

**Species**  
Select All  
ALLIUM CEPA (ONION)  
ASPERGILLUS FLAVUS  
BACILLUS SUBTILIS  
BACTERIOPHAGE

**Strain**  
Select All

**Cells/Cell Lines**  
Select All  
OTHER CELLS IN VIT  
HEPATOCYTE  
OTHER CELLS IN VIV  
LYMPHOCYTE

**Metabolic Activation**  
Select All  
unspecified  
Present  
Absent

**Study Call**  
Select All  
Negative  
Positive

**Test Call**  
Select All  
Negative  
Positive

Buttons: Add Query, Add & Close, Close

**Footer:**  
Molecular Networks Inspiring Chemical Discovery  
Designed and Developed by Altamira LLC

# Searching the Database

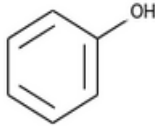
- Combined chemistry and toxicity data

**COSMOS** Database Search Home ▾ About Logout (fabian.steinmetz)

Query Definitions: Chemistry

Names Identifiers CAS Registry Number Structure: Sketch molecule Structure: SMILES

Edit molecule Clear



SMILES: C1=CC=CC=C1O

Options  
☐ exact ☐ partial ☒ similar  similar

Query Definitions: Toxicity Studies

+ clear

Active Queries

- Subchronic
  - Species: rat
  - Strain: Select All
  - Sex: Select All
  - Route of Exposure: Oral
  - Site: KIDNEY
- Subchronic
  - Species: rat
  - Strain: Select All
  - Sex: Select All
  - Route of Exposure: Oral
  - Site: LIVER
- Subchronic
  - Species: rat
  - Strain: Select All
  - Sex: Select All
  - Route of Exposure: Oral

Search

# Demonstration

- Search a substance/list of substances
  - > search by name, CAS number, draw structure, SMILES
  - > examples for exact, partial, similar search
- “find information on substances”
  - > what kind of studies, use function, link to CosIng, SCCS opinions etc, what effects, dose levels etc
- Search for toxicity
  - > find substances for certain effects
- Combined chemistry and toxicity search
  - > e.g. for CAS nos. find all subchronic studies with liver effects

# Exercise: Database Search

- For compounds with CAS 59820-63-2 and 80062-31-3 find all available subchronic studies in which effects in liver were observed.



# Summary COSMOS DB

- COSMOS DB is an open access database:
  - Cosmetics Inventory
  - Toxicity information
- Uses of COSMOS DB:
  - To support risk assessment
  - To capture and analyse knowledge e.g. grouping, TTC
- High standards of QA / QC and curation
- Support for 21<sup>st</sup> Century Toxicology and AOP development
- Capacity for future growth

More information:

<http://www.cosmostox.eu/what/COSMOSdb>

- Database webinar recording
- Short tutorial

# Acknowledgements



The funding from the European Community's  
Seventh Framework Program (FP7/2007-2013)  
COSMOS Project under grant agreement n° 266835  
and from Cosmetics Europe

