

The Data Available

Christine Chichester
Swiss Institute of Bioinformatics



Using the Power of Open PHACTS

Open PHACTS 4th Community Workshop, London, 22-23 April 2013



Scientific Focus

- ✦ Pharmacological Sciences
 - Bioactive drug-like small molecules
 - Pharmacological targets: proteins and protein complexes
- ✦ Biological Sciences
 - Genes
 - Proteins
 - Protein isoforms
 - Pathways
 - Diseases
- ✦ Chemistry
 - Salt forms
 - Racemic mixtures
 - Stereospecific forms
 - Isotopically labeled
 - Tautomeric forms
 - Mixtures

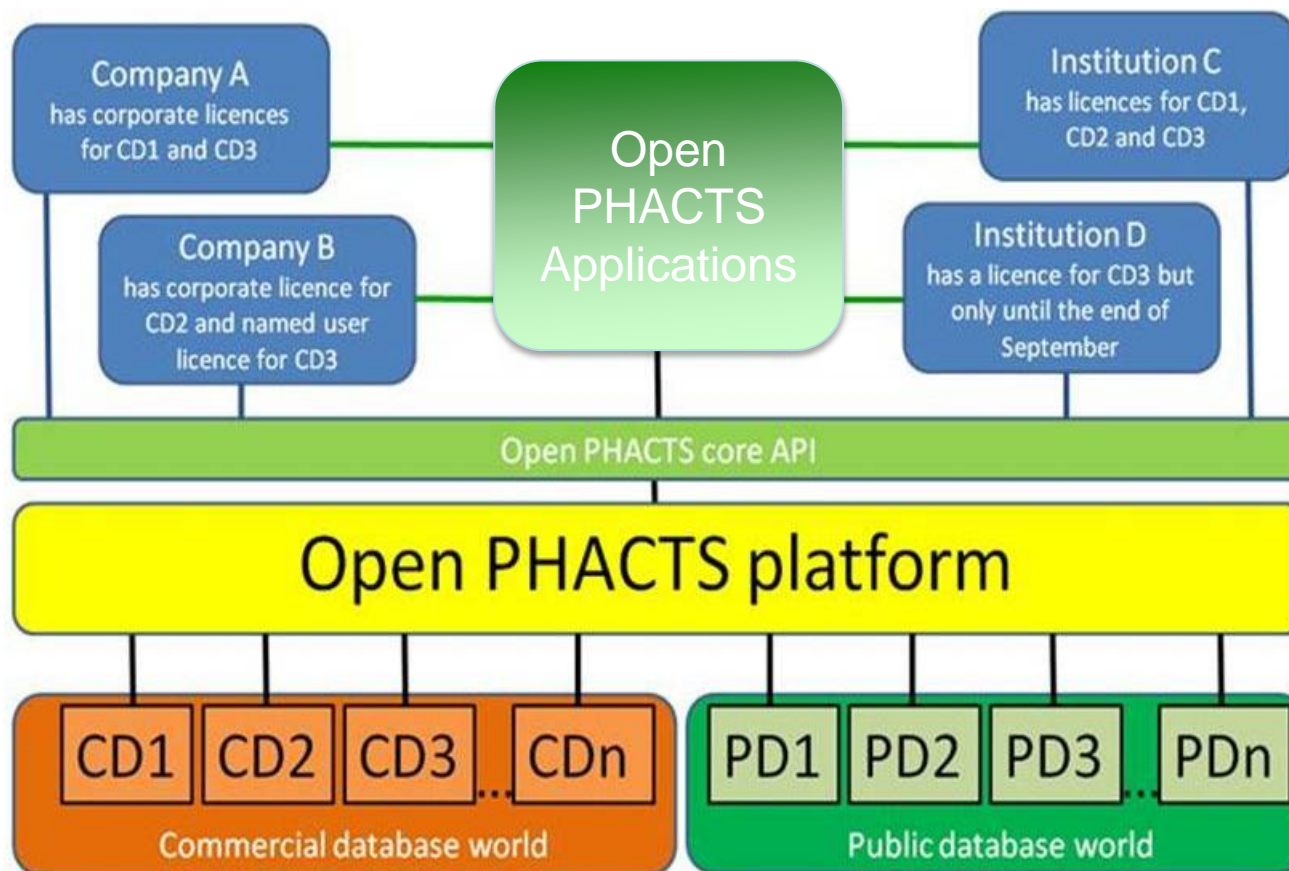


Present Content

<u>Source</u>	<u>Initial Records</u>	<u>Triples</u>	<u>Properties</u>
ChEMBL	1,149,792 ~1,091,462 cmpds ~8845 targets	146,079,194	17 cmpds 13 targets
DrugBank	19,628 ~14,000 drugs ~5000 targets	517,584	74
UniProt	536,789	156,569,764	78
ENZYME	6,187	73,838	2
ChEBI	35,584	905,189	2
GO/GOA	38,137	24,574,774	42
ChemSpider/ACD	1,194,437	161,336,857	22 ACD, 4 CS
ConceptWiki	2,828,966	3,739,884	1



Commercial Data



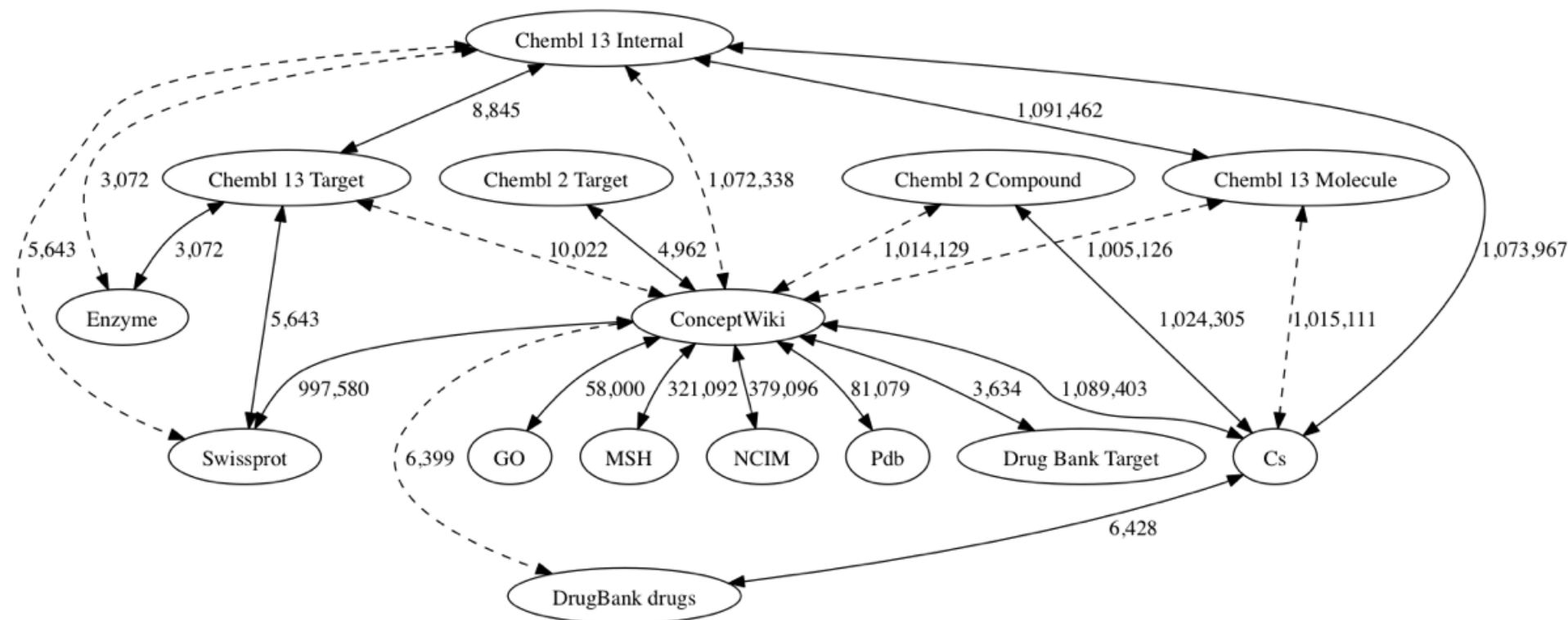


Future Content

- ✦ neXtProt
- ✦ Wikipathways
- ✦ DisGenNet
- ✦ Literature content from text mining
- ✦ Nanopublications
 - neXtProt PTM and gene expression nanopublications
 - Chemistry journals (RSC) nanopublications
 - Drugbank nanopublications
 - Manually curated nanopublications

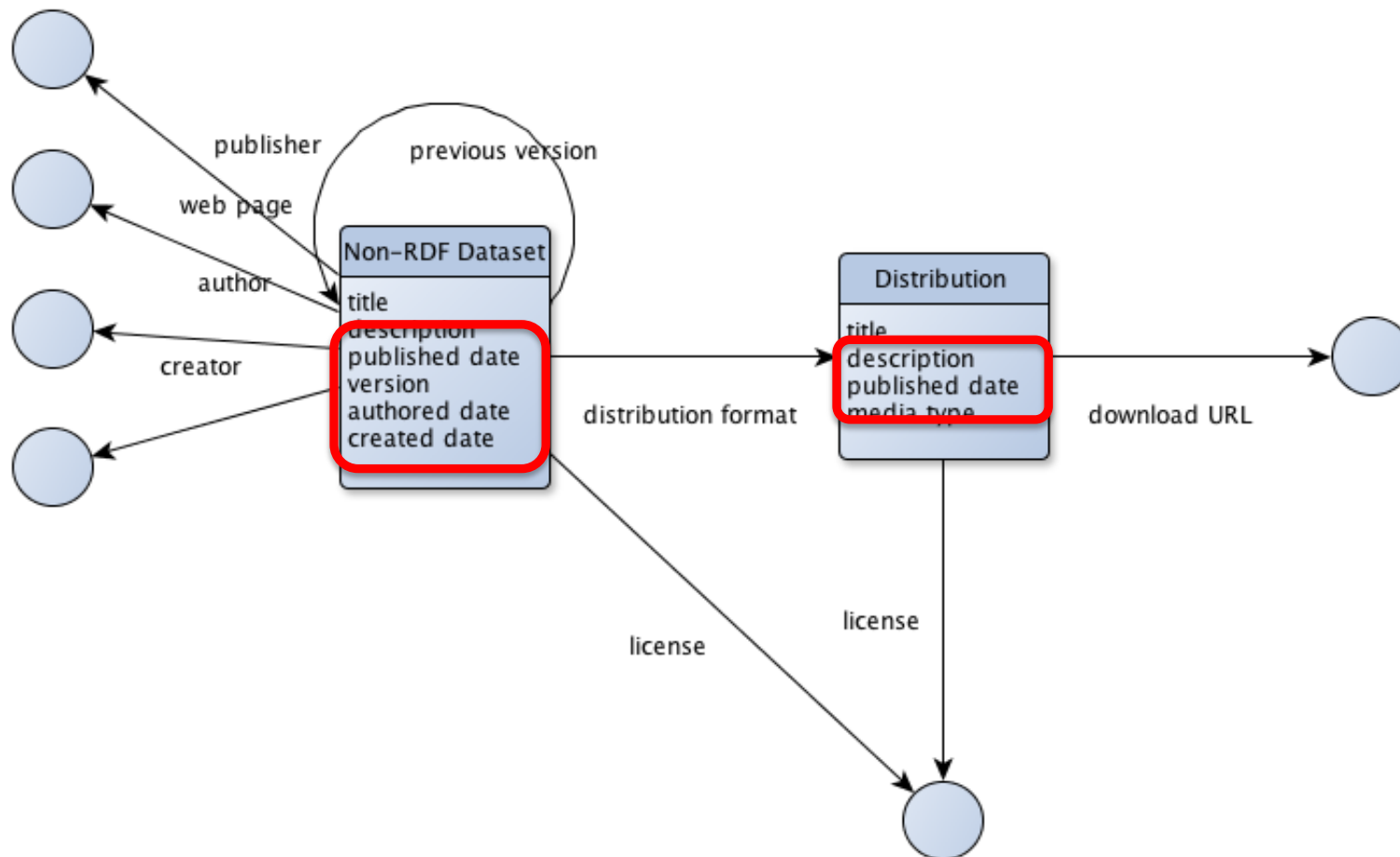


Mappings





Version tracking





Mapping ChEMBL targets classes to concepts

Enzyme

Enzyme

Enzyme Kinase

Kinase

Enzyme Kinase Protein
Kinase

Protein Kinase

Enzyme Kinase Protein
Kinase Ser_Thr

Serine/threonine-protein kinase

Enzyme Kinase Protein
Kinase Ser_Thr Agc

AGC kinase group

Enzyme Kinase Protein
Kinase Ser_Thr Agc Akt

protein kinase B (Akt)

Enzyme Kinase Protein
Kinase Ser_Thr Agc Ndr

nuclear Dbf2-related kinase
(Ndr)

Enzyme Kinase Protein
Kinase Ser_Thr Agc Pka

Protein Kinase A

Enzyme Kinase Protein
Kinase Ser_Thr Agc Pkc

protein kinase C

Enzyme Kinase Protein
Kinase Ser_Thr Agc Pkg

Protein Kinase G

Enzyme Kinase Protein
Kinase Ser_Thr Agc Pkn

protein kinase N



Using the Power of Open PHACTS, London, 22-23 April 2013

Hierarchies

- ❖ GeneOntology
- ❖ Chemical
- ❖ ENZYME
- ❖ Pathway
- ❖ ChEMBL

Enzyme (3727)

Kinase (689)

Protein kinase

Serine

Tyrosine

Serine

Histidine

Guanine

Endonuclease

Serine

Protease (4)

Phosphodiesterase

Phosphatase (70)

Cytochrome P450 (57)

Aminoacyltransferase (1)

Reductase (1)

Unclassified (1114)

Membrane receptor (747)

Ion channel (519)

Transporter (162)

Transcription Factor (148)

Cytosolic other (119)

Secreted (59)

Structural (33)

Surface antigen (26)

Membrane other (22)

Adhesion (15)

Nuclear other (15)

I classic metabolic pathway + A

I disease pathway + A

I drug pathway + A

I regulatory pathway + A

I signaling pathway + A

cyclase
cleave



Concept based

- ✦ Chemical compounds
- ✦ Biological targets
- ✦ Pathways
- ✦ Diseases



Using the Power of Open PHACTS, London, 22-23 April 2013

Semantic types

Physical Object	Activity		Functional Concept
Organism	Behavior		Body System
Plant	Social Behavior		Spatial Concept
Fungus	Individual Behavior		Body Space or Junction
Virus	Daily or Recreational Activity		Body Location or Region
Bacterium	Occupational Activity		Molecular Sequence
Archaeon	Health Care Activity		Nucleotide Sequence
Eukaryote	Laboratory Procedure		Amino Acid Sequence
Animal	Diagnostic Procedure		Carbohydrate Sequence
Vertebrate	Therapeutic or Preventive Procedure		Geographic Area
Amphibian	Research Activity		Finding
Bird	Molecular Biology Research Technique		Laboratory or Test Result
Fish	Governmental or Regulatory Activity		Sign or Symptom
Reptile	Educational Activity		Organism Attribute
Mammal	Machine Activity		Clinical Attribute
Human	Phenomenon or Process		Intellectual Product
Anatomical Structure	Human-caused Phenomenon or Process		Classification
Embryonic Structure	Environmental Effect of Humans		Regulation or Law
Anatomical Abnormality	Natural Phenomenon or Process		Language
Congenital Abnormality	Biologic Function		Occupation or Discipline
Acquired Abnormality	Physiologic Function		Biomedical Occupation or Discipline
Fully Formed Anatomical Structure	Organism Function		Organization
Body Part, Organ, or Organ Component	Mental Process		Health Care Related Organization
Tissue	Organ or Tissue Function		Professional Society
Cell	Cell Function		Self-help or Relief Organization
Cell Component	Molecular Function		Group Attribute
Gene or Genome	Genetic Function		Group
Manufactured Object	Pathologic Function		Professional or Occupational Group
Medical Device	Disease or Syndrome		Population Group
Drug Delivery Device	Mental or Behavioral Dysfunction		Family Group
Research Device	Neoplastic Process		Age Group
Clinical Drug	Cell or Molecular Dysfunction		Patient or Disabled Group
	Experimental Model of Disease		
	Injury or Poisoning		



Concept 60915889-bbea-4c10-a810-7dfd6eb05168



view / [edit](#)

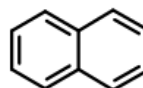
Mothballs

by ConceptWiki (all)

Mothballs [Naphthalene](#) [Naphthaline](#) [Moth balls](#) [214-552-7](#) [91-20-3](#)

Naphthalene

by ChemSpider



Chemical Viewed Structurally

Naphthalene [\(2H8\)naphthalene](#) [NPY](#) [Naphthalene Zone Refined](#)
(number of passes:23) [72931-45-4](#) [Moth balls](#) [naftalina](#) [tar](#)
[camphor](#) [NAPHTHALENE PURE](#) [68412-25-9](#) [1146-65-2](#) [Naphthalen](#) [Camphor tar](#)

naphthalene

by UMLS

Organic Chemical

naphthalene

Notations

Unified Medical Language System C0027375

Medical Subject Headings C031721

associated labels

[\(2H8\)naphthalene](#) [1146-65-2](#) [214-552-7](#) [68412-25-9](#)
[72931-45-4](#) [91-20-3](#) [Camphor tar](#) [Moth Flakes](#) [Moth balls](#) [Mothballs](#) [NAPHTHALENE, PURE](#) [NPY](#)
[Naphthalen](#) [Naphthalene](#)
[\(1mg/ml in Methanol\) \[for Water Analysis\]](#) [Naphthalene Zone Refined](#) (number of passes:23) [Naphthaline](#) [naftalina](#)
[naphthalene](#) [tar](#) [camphor](#)

9 most recent searches

[naphthalene, 9,10-anthracenediol, 9,10-anthracendiol,](#)
[anthracene-9, 10-diol , 1h-thieno\[3,4-g\]indazole, 4,5-](#)
[dihydro-1,6,8-trimethyl-, 1,3-dibenzyltetrahydro-1h-](#)



Different views on content

- ✦ Genes/Proteins
- ✦ Racemic mixtures/ Enantiomers
- ✦ Species specific protein/ All homologous proteins
- ✦ Single target proteins/ Protein complexes
- ✦ Compound drugs/ Individual chemicals
- ✦ Validated/ Nonvalidated data



Public domain databases

- ❖ Many databases have messy content
- ❖ Non-curated databases proliferate errors
- ❖ We source and deposit data between databases
- ❖ Sources of original errors are difficult to determine
- ❖ Curation is time consuming and challenging



Community Annotations





Identify records to deprecate

<u>23412</u>		5/1/2008 1:02:27 PM	Extreme	New	<u>This structure should be deprecated. The structure is abysmal</u>	Antony Williams antony.williams@chemspider.com
<u>16352795</u>		4/29/2008 3:42:04 PM	Normal	New	<u>disodium salt or anthracene-9,10-diol ???</u>	kolshorn@uni-mainz.de
<u>5558</u>		4/24/2008 2:52:04 AM	High	New	<u>I look for 2-Butenedioic acid (2Z)-, homopolymer (RN 26099-09-2), but I get 50-37-3. It does not seem alright. Searching with key word of 26099-0902, no results can be found. But in ChemIDplus, it does exist.</u>	alxu@sina.com





Open PHACTS
Open Pharmacological Space



Using the Power of Open PHACTS, London, 22-23 April 2013

Names and Synonyms

Synonym curation

selection | Deselect all | Invert selection |

Validated by Experts, Validated by Users, Non-Validated, Removed by Users, Redirected by Users, Redirect Approved by Experts

Identifiers

Names and Synonyms

Validated by Experts, Validated by Users, Non-Validated, Removed by Users, Redirected by Users, Redirect Approved by Experts

1H-thieno[3,4-d]imidazole-4-pentanoic acid, hexahydro-2-oxo-, (3aR,4S,6aR)-

5-[(3aR,4S,6aR)-2-Oxohexahydro-1H-thieno[3,4-d]imidazol-4-yl]pentanoic acid

58-85-5 [RN]

D(+)-Biotin >99.5%

VITAMIN H

Normal
Normal
Confirm
Confirmed
Approve
Approved
Common Name
Redirect
Redirected
Redirect Approved



ChemSpider
The free chemical database



Concept 60915889-bbea-4c10-a810-7dfd6eb05168

view / [edit](#)

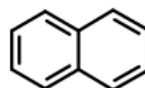
Mothballs

by ConceptWiki (all)

Mothballs [Naphthalene](#) [Naphthaline](#) [Moth balls](#) [214-552-7](#) [91-20-3](#)

Naphthalene

by ChemSpider



Chemical Viewed Structurally

Naphthalene [\(2H8\)naphthalene](#) [NPY](#) [Naphthalene Zone Refined](#)
(number of passes:23) [72931-45-4](#) [Moth balls](#) [naftalina](#) [tar](#)
[camphor](#) [NAPHTHALENE PURE](#) [68412-25-9](#) [1146-65-2](#) [Naphthalen](#) [Camphor tar](#)

naphthalene

by UMLS

Organic Chemical

naphthalene

Notations

Unified Medical Language System C0027375

Medical Subject Headings C031721

associated labels

[\(2H8\)naphthalene](#) [1146-65-2](#) [214-552-7](#) [68412-25-9](#)
[72931-45-4](#) [91-20-3](#) [Camphor tar](#) [Moth Flakes](#) [Moth balls](#) [Mothballs](#) [NAPHTHALENE, PURE](#) [NPY](#)
[Naphthalen](#) [Naphthalene](#)
[\(1mg/ml in Methanol\) \[for Water Analysis\]](#) [Naphthalene Zone Refined](#) (number of passes:23) [Naphthaline](#) [naftalina](#)
[naphthalene](#) [tar](#) [camphor](#)

9 most recent searches

[naphthalene, 9,10-anthracenediol, 9,10-anthracendiol,](#)
[anthracene-9, 10-diol , 1h-thieno\[3,4-g\]indazole, 4,5-](#)
[dihydro-1,6,8-trimethyl-, 1,3-dibenzyltetrahydro-1h-](#)



Open PHACTS

Open Pharmacological Space



Using the Power of Open PHACTS, London, 22-23 April 2013

existing labels

- Preferred labels cannot be deleted. For each language, first promote another label to "Preferred" and then delete the old preferred label.
- Only one "Preferred" label is allowed and required per language.

preferred	language	label	delete
<input type="button" value="preferred"/>	en	Naphthalene	<input type="button" value="delete"/>
<input type="button" value="preferred"/>	en	Naphthaline	<input type="button" value="delete"/>
<input type="button" value="preferred"/>	en	Moth balls	<input type="button" value="delete"/>
<input type="button" value="preferred"/>	en	214-552-7	<input type="button" value="delete"/>
<input type="button" value="preferred"/>	en	91-20-3	<input type="button" value="delete"/>
<input checked="" type="button" value="preferred"/>	en	Mothballs	<input type="button" value="delete"/>



Open PHACTS

Open Pharmacological Space



Using the Power of Open PHACTS, London, 22-23 April 2013

"Naphthalene" was set as the preferred label



add a label

en

Add

existing labels

- Preferred labels cannot be deleted. For each language, first promote another label to "Preferred" and then delete the old preferred label.
- Only one "Preferred" label is allowed and required per language.

preferred language label

delete

preferred	en	Mothballs	delete
preferred	en	Naphthaline	delete
preferred	en	Moth balls	delete
preferred	en	Naphthalene	delete
preferred	en	214-552-7	delete
preferred	en	91-20-3	delete



Concept 60915889-bbea-4c10-a810-7dfd6eb05168

view / [edit](#)

Naphthalene

by ConceptWiki (all)

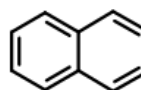
Naphthalene Mothballs Naphthaline Moth balls 214-552-7 91-20-3

Naphthalene

by ChemSpider

Chemical Viewed Structurally

Naphthalene (2H8)naphthalene NPY Naphthalene Zone Refined (number of passes:23) 72931-45-4 Moth balls naftalina tar camphor NAPHTHALENE, PURE 68412-25-9 1146-65-2 Naphthalen Camphor tar Naphthalene (1mg/ml in Methanol) [for Water Analysis] Moth Flakes Mothballs 202-049-5



naphthalene

by UMLS

Organic Chemical

naphthalene

Notations

Unified Medical Language System C0027375

Medical Subject Headings C031721

associated labels

(2H8)naphthalene 1146-65-2 214-552-7 68412-25-9 72931-45-4 **91-20-3** Camphor tar Moth Flakes **Moth balls Mothballs** NAPHTHALENE, PURE NPY Naphthalen **Naphthalene** Naphthalene (1mg/ml in Methanol) [for Water Analysis] Naphthalene Zone Refined (number of passes:23) Naphthaline naftalina naphthalene tar camphor

9 most recent searches Q

naphthalene, 9,10-anthracenediol, 9,10-anthracendiol, anthracene-9, 10-diol , 1h-thieno[3,4-g]indazole, 4,5-dihydro-1,6,8-trimethyl-, 1,3-dibenzyltetrahydro-1h-thieno[3,4-d]imidazol-2(3h)-one, vitamin d, vitamin H, vitamin D



Communication with data providers

- ❖ ChemSpider
- ❖ ChEMBL
- ❖ neXtProt
- ❖ Pathway Ontology
- ❖ Wikipathways
- ❖ Bio2RDF
- ❖ Commercial data providers



Thank you for your attention