Open PHACTS
Open Pharmacological Space

Scientific Lenses over Linked Chemistry Data using BridgeDb and the Open PHACTS Chemistry Registration System

Colin Batchelor	***Royal Society of Chemistry***
Christian Brenninkmeijer	*University of Manchester*
Chris Evelo	*Maastricht University*
Carole Goble	*University of Manchester*
Alasdair Gray	*Heriot-Watt University*
Ken Karapetyan	*Royal Society of Chemistry*
Valery Tkachenko	*Royal Society of Chemistry*
Egon Willighagen	*Maastricht University*

Open PHACTS
Open Pharmacological Space

Overview

- Background to Open PHACTS
- Motivation
- Lenses
- How the Open PHACTS Chemistry Registration Service helps

Open PHACTS
Open Pharmacological Space

Open PHACTS provides an ***“integrated platform of publicly available pharmacological and physicochemical data”***

ChemSpider Search and share chemistry
DrugBank Open Data Drug & Drug Target Database
UniProt
ChEMBL
swissprot
ChEBI
Concept Wiki
GENEONTOLOGY Unifying Biology
WikiPathways Pathways for the People

Accessible via: Open PHACTS Discovery Platform

- Application programming interface (API)
- Third-party applications built to use the API

dev.openphacts.org **www.openphactsfoundation.org**

Open PHACTS
Open Pharmacological Space

How does it all work?

The underlying language of Open PHACTS is **RDF**.

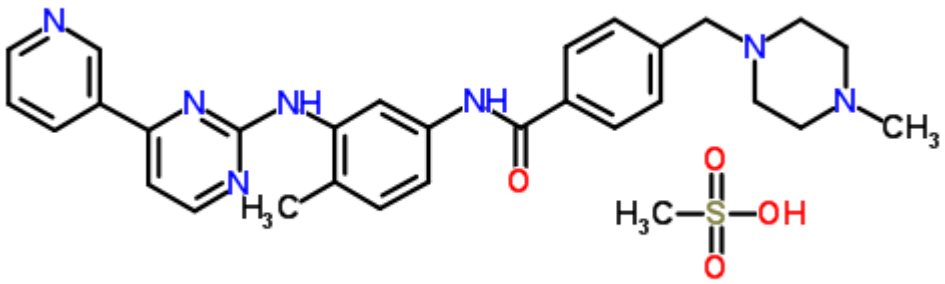
There are few constraints as such, but we provide **guidelines** for which classes of identifier to use, accounts of **best practice**, and attempt to deal with RDF as we find it.

This RDF goes into the **data cache** and we access the results through user interfaces built on **RESTful** web services offering **JSON**, **RDF, HTML** and **TSV.**

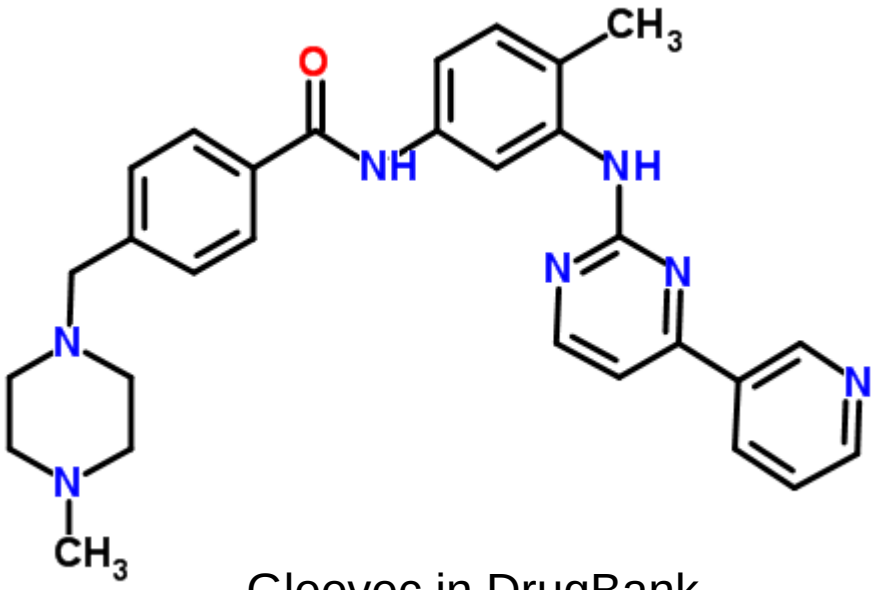

by Philip Chapman-Bell (https://www.flickr.com/photos/oschene/)
Magnifying Lens Price Guide:
2” - $1.99 2.5” - $2.99 3” - $3.99 3.5” - $4.99

Open PHACTS
Open Pharmacological Space

Motivation



Gleevec in ChemSpider and ChEMBL
(imatinib mesylate)



Gleevec in DrugBank
(actually imatinib)

How is this a semantic web problem? Why can't people just be clear?

People may be working with faulty data.

Salts, say, may make little difference to the effects of an active ingredient.

People may assume a one-to-one mapping between a gene and the gene product (protein, ncRNA) that it codes for.

What's in a lens?

Identifier

Title (`dct:title`)

Description (`dct:description`)
Documentation link (`dcat:landingPage`)
Creator (`pav:createdBy`)

Timestamp (`pav:createdOn`)

Equivalence rules (`bdb:linksetJustification`)

Equivalence rules

The BridgeDB vocabulary adds metadata that provides a justification for treating two URIs alike, thus allowing the researcher to determine whether their circumstances fit.

owl:sameAs ≤ skos:exactMatch ≤ skos:closeMatch ≤ rdfs:seeAlso

The ChEBI and CHEMINF ontologies provide a rich set of relations (many of which developed for this project) to relate one molecule to another.

Open PHACTS
Open Pharmacological Space

ChEBI (http://www.ebi.ac.uk/chebi**)**

has part
is tautomer of

CHEMINF (http:**//**code.google.com/p/semanticchemistry/**)**

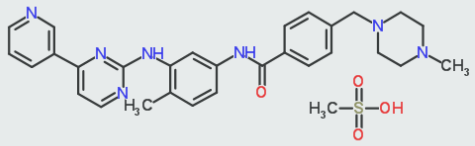
has component with uncharged counterpart
has counterpart molecular entity
- has normalized counterpart
  - has OPS normalized counterpart
  - has PubChem normalized counterpart
- has uncharged counterpart
- similar to
  - similar to by PubChem 2D similarity algorithm
  - similar to by PubChem 3D similarity algorithm

has same connectivity as
- is isotopologue of
- is stereoisomer of

subClassOf (standard relation in RDF)
- has isotopically unspecified parent
- has stereoundefined parent

Open PHACTS
Open Pharmacological Space

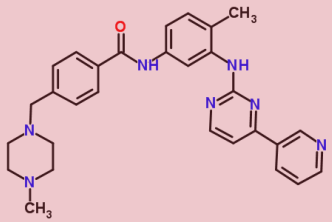
ChemSpider
Search and share chemistry



Link: skos:closeMatch
Reason: non-salt form

Link: skos:exactMatch
Reason: drug name

DrugBank
Open Data Drug & Drug Target Database



Open PHACTS
Open Pharmacological Space

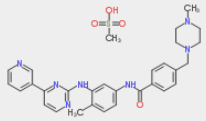
Strict

Relaxed

Analysing

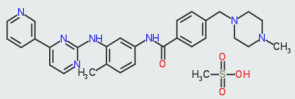
Browsing

ChEMBL

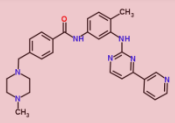


skos:exactMatch
(InChI)

ChemSpider
Search and share chemistry



DrugBank
Open Data Drug & Drug Target Database



Open PHACTS
Open Pharmacological Space

Strict

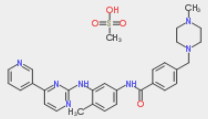
Relaxed

Analysing

Exploring

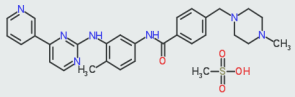
skos:closeMatch
(Drug Name)

ChEMBL



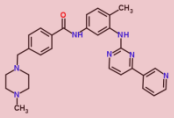
skos:exactMatch
(InChI)

ChemSpider
Search and share chemistry



skos:closeMatch
(Drug Name)

DrugBank
Open Data Drug & Drug Target Database



Open PHACTS
Open Pharmacological Space

What does the Open PHACTS Chemistry Registration System do?

Takes in structures from ChEMBL, ChEBI, DrugBank, PDB, Thomson Reuters.

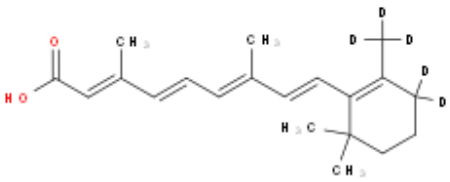
Normalizes structures according to rules based on FDA guidelines.

Generates counterpart molecules: without charge, fragments

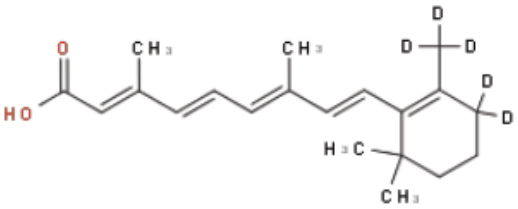
Open PHACTS

ROYAL SOCIETY OF CHEMISTRY

Compound Record

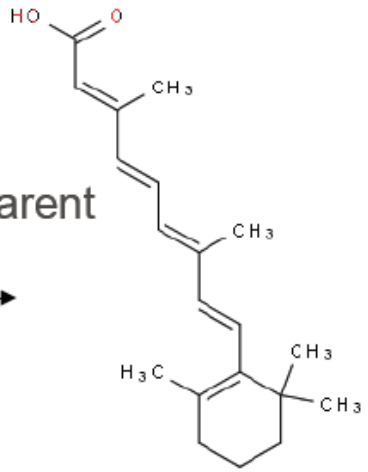


ID	45975
Virtual	No
Molecular Formula	$C_{20}H_{28}O_2$
Monoisotopic Mass	305.240326 Da
Molecular Weight	305.465942 Da
SMILES	[2H]C([2H])([2H])C1=C(/C=C/C(/C)=C/C=C/C(/C)=C/C(O)=O)C(C)(C-)CCC1([2H])[2H]
Non Std. InChI	InChI=1/C20H28O2/c1-15(8-6-9-16(2)14-19(21)22)11-12-18-17(3)-10-7-13-20(18,4)5/h6,8-9,11-12,14H,7,10,13H2,1-5H3,(H,21,22)/b9-6+,12-11+,15-8+,16-14+/i3D3,10D2/f/h21H
Non Std. InChIKey	SHGAZHPCJJPHSC-HPEAYTSRNA-N
Std. InChI	InChI=1S/C20H28O2/c1-15(8-6-9-16(2)14-19(21)22)11-12-18-17(3-)10-7-13-20(18,4)5/h6,8-9,11-12,14H,7,10,13H2,1-5H3,(H,21,22)/b9-6+,12-11+,15-8+,16-14+/i3D3,10D2
Std. InChIKey	SHGAZHPCJJPHSC-GOWYKSNKSA-N
ChemSpider ID	23937116



http://ops.rsc.org/OPS45975

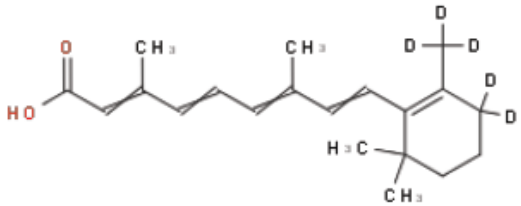
has_isotopically_unspecified_parent
[CHEMINF:000459]



http://ops.rsc.org/OPS45978

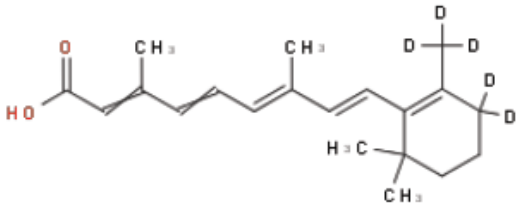
has OPS normalized counterpart
[CHEMINF:000458]

has_stereoundefined_parent
[CHEMINF:000456]

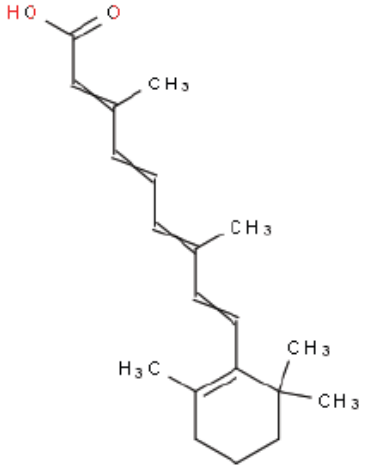


http://ops.rsc.org/OPS45981

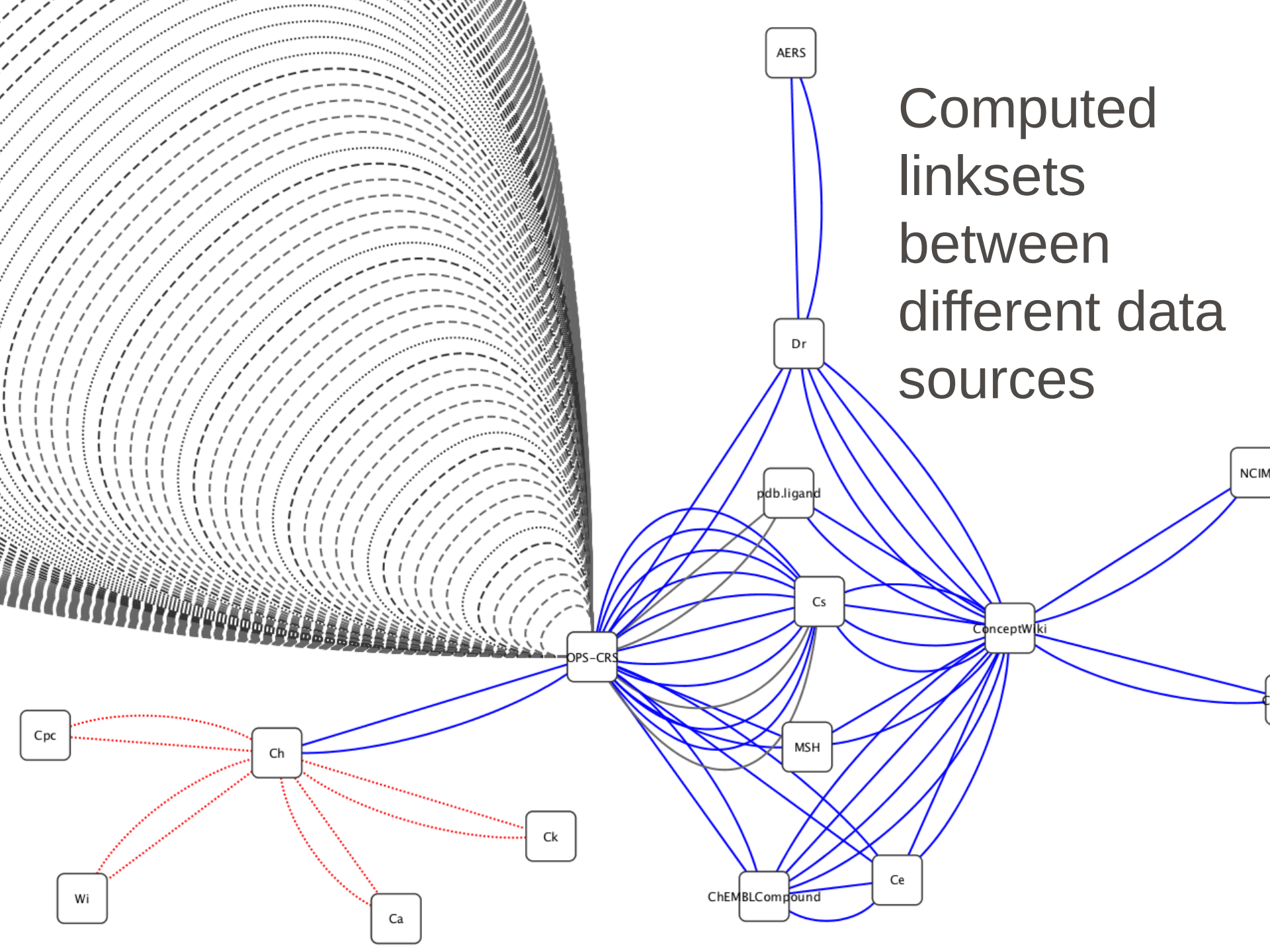
is_tautomer_of
[chebi:is_tautomer_of]



http://ops.rsc.org/OPS45987



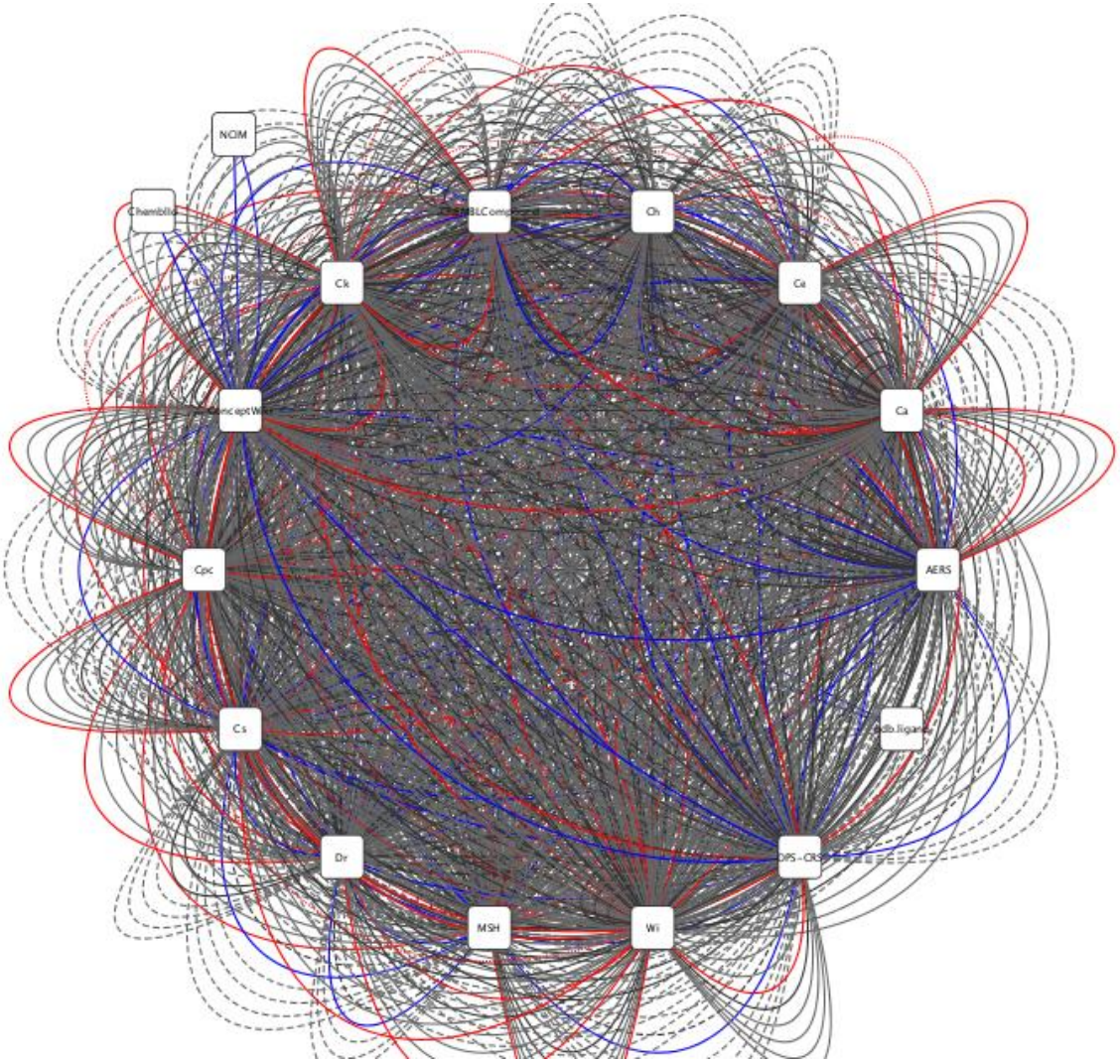
http://ops.rsc.org/OPS45991



Computed linksets between different data sources

Open PHACTS

Open Pharmacological Space



Where do mappings come from?

1. Chemical Registration Service
2. BridgeDb
    1. Metabolites: HMDB, ChEBI
    2. Genes / Proteins: Ensembl (*)

ChemSpider
Search and share chemistry

BridgeDb

(*) Andra Waagmeester and Alex Pico (WikiPathways).

Open PHACTS
Open Pharmacological Space

Those chemistry lenses in full

OPS Default
Chemistry Tautomer
Chemistry Parent→Child
ConceptWiki Gene
Chemistry Parent→Child plus Tautomer
Chemistry Child→Parent plus Tautomer
Chemistry Child→Parent
Chemistry Strict

Source:
http://openphacts.cs.man.ac.uk:9092/QueryExpander/SourceTargetInfos

Open PHACTS
Open Pharmacological Space

How do I access all this?

Open PHACTS
Open Pharmacological Space

The Open PHACTS API dev.openphacts.org

Launched in January 2013 – Over 16 million queries for data

Chemical Structure Exact Search	/structure/exact	GET
InchiKey to URL	/structure	GET
Inchi to URL	/structure	GET
Chemical Structure Similarity Search	/structure/similarity	GET
SMILES to URL	/structure	GET
Chemical Structure Substructure Search	/structure/substructure	GET
Map URL	/mapUri	GET

Open PHACTS
Open Pharmacological Space

Open PHACTS Discovery Platform

Open PHACTS API

dev.openphacts.org

Applications can query the pharmacological data within Open PHACTS

- Compound-protein interactions
- Physicochemical properties
- Gene information
- Biological pathways

Open PHACTS Explorer
Web based searching interface
explorer.openphacts.org

Workflow tools
Pipeline Pilot, KNIME, R

Open PHACTS applications
External bespoke applications using the Open PHACTS API.

chembionavigator.org

pharmatrek.org

Open PHACTS
Open Pharmacological Space

Open PHACTS Explorer

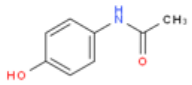
jupiter.cs.man.ac.uk/compounds?uri=http%3A%2F%2Fwww.conceptwiki.org%2Fconcept

Google

Open PHACTS Explorer Alpha | paracetamol | Search | Browse by Ontology

Home / paracetamol / Compound Information

Provenance On Off



paracetamol

$C_8H_9NO_2$

Pharmacology (2074)

Structure

Draw Molecule

Analgesic antipyretic derivative of acetanilide. It has weak anti-inflammatory properties and is used as a common analgesic, but may cause liver, blood cell, and kidney damage. [PubChem]

AlogP
1.082

H-Bond Acceptors
3

H-Bond Donors
2

Mol Weight
151.163

ChemSpider ID	OPS1019246
SMILES	CC(=O)NC1=CC=C(C=C1)O
Standard InChI	InChI=1S/C8H9NO2/c1-6(10)9-7-2-4-8(11)5-3-7/h2-5,11H,1H3,(H,9,10)
Standard InChIKey	RZVAJINKPMORJF-UHFFFAOYSA-N
Protein Binding	25%
Toxicity	Oral, mouse: LD₅₀ = 338 mg/kg; Oral, rat: LD₅₀ = 1944 mg/kg. Acetaminophen is metabolized primarily in the liver, where most of it is converted to inactive compounds by conjugation with sulfate and glucuronide, and then excreted by the kidneys. Only a small portion is metabolized via the hepatic cytochrome P450 enzyme system. The toxic effects of acetaminophen are due to a minor alkylating metabolite (N-acetyl-p-benzo-quinone imine), not acetaminophen itself nor any of the major

Open PHACTS
Open Pharmacological Space

Who is involved?

USC Universidade de Santiago de Compostela
universitätbonn
AstraZeneca
gsk GlaxoSmithKline
Universität Hamburg DER FORSCHUNG | DER LEHRE | DER BILDUNG
BioSolveIT
ESTEVE
NOVARTIS
MERCK
Lundbeck
universität wien
DTU
Lilly
nbic
SIB Swiss Institute of Bioinformatics
connected discovery
ROYAL SOCIETY OF CHEMISTRY
VU UNIVERSITY AMSTERDAM
EMBL-EBI
janssen
OPENLINK SOFTWARE Making Technology Work For You
AQnowledge Semantics for Science
MANCHESTER 1824 The University of Manchester
Maastricht University
Parc de Salut Barcelona MAR
Pfizer
LUMC Leids Universitair Medisch Centrum
cnio
The Open PHACTS Foundation

Innovative Medicines Initiative
efpia European Federation of Pharmaceutical Industries and Associations

29 partners

Open PHACTS
Open Pharmacological Space

The Open PHACTS Foundation - Established May 2013

A not-for-profit, member-owned successor organisation, based in UK:

- Acts a **unique forum** for the scientific community, bringing EFPIA partners together with academia and SMEs.
- Ensures the **sustainability** of the Open PHACTS infrastructure, and provides technical support
- A hub of **relevant scientific research** and development
- Full members **nominate and vote for Directors**, influencing the Foundation's development and strategy

www.openphactsfoundation.org

Acknowledgements

The Open PHACTS project has received support from the Innovative Medicines Initiative Joint Undertaking under grant agreement n° 115191, resources of which are composed of financial contribution from the European Unions's Seventh Framework Programme (FP7/2007-2013) and EFPIA companies' in kind contribution.

Jon Steele (Royal Society of Chemistry) and many other participants in Open PHACTS.

Open PHACTS
Open Pharmacological Space

Any questions?

More information about Open PHACTS

Project: **www.openphacts.org**
Foundation: **www.openphactsfoundation.org**
Twitter: **@open_phacts**
E-mail: **support@openphacts.org**

Access the data:

API: **dev.openphacts.org**
Explorer: **explorer.openphacts.org**

batchelorc@rsc.org
@documentvector

Open PHACTS