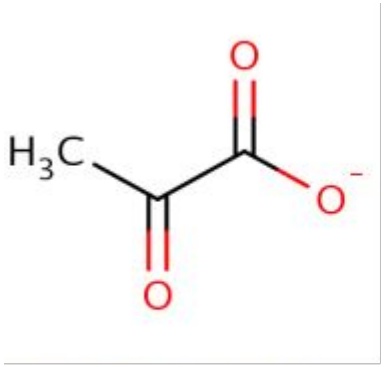
Cheminformatics to improve Wikidata on chemical compounds

Egon Willighagen (Q20895241)
WikdiataCon 2019
2019-10-26, Berlin/DE
ORCID: 0000-0001-7542-0286
@egonwillighagen
#WikidataCon

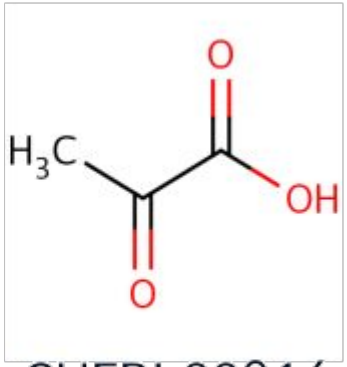
Maastricht University

Chemistry in metabolic pathways

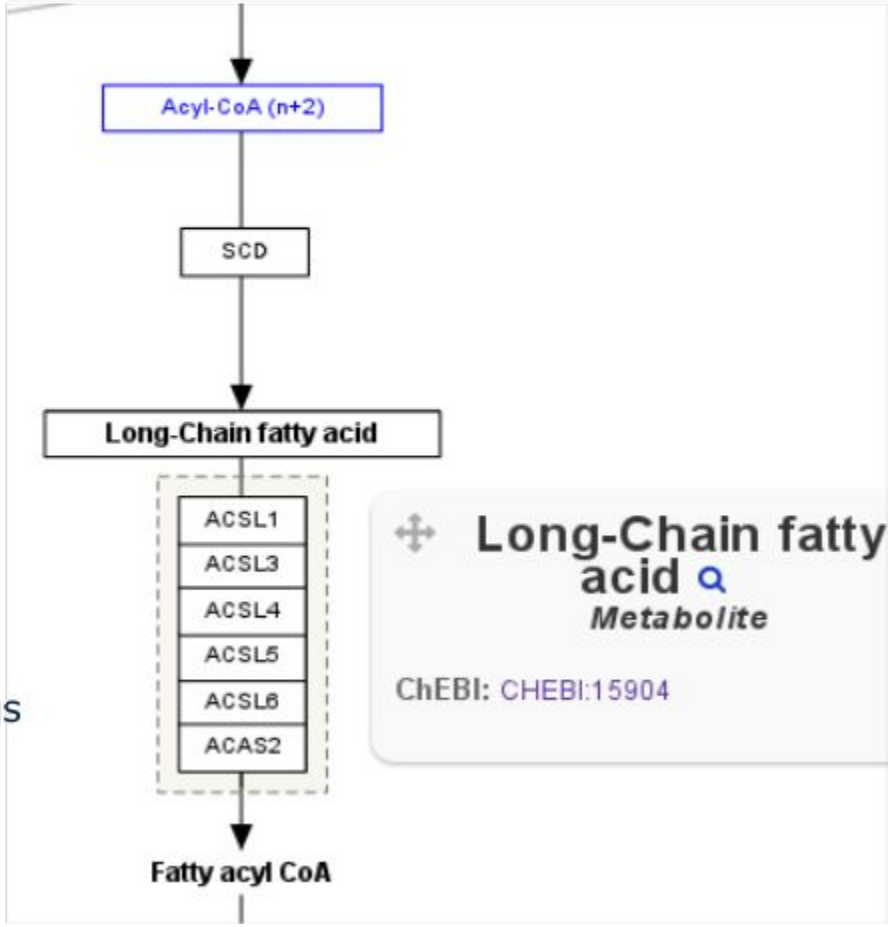
CHEBI:15361 (Pyruvate) -> Ce:CHEBI:32816 (conjugate) -> Ck:C00022 -> [WP2456 HIF1A and PPARG regulation of glycolysis, WP2453 TCA Cycle and PDHc]



CHEBI:15361



CHEBI:32816



Brenninkmeijer, CYA, et al. "Scientific Lenses over Linked Data: An approach to support task specific views of the data. A vision." Proceedings of 2nd International Workshop on Linked Science. 2012.

So, what IDs are used in WikiPathways?

2017

datasource	numberEntries
ChEBI	1923
HMDB	623
CAS	299
KEGG Compound	251
PubChem-compound	245
Chemspider	174
PubChem-substance	33
LIPID MAPS	10
Reactome	4
Wikidata	3
ChEMBL compound	2

+ Reactome

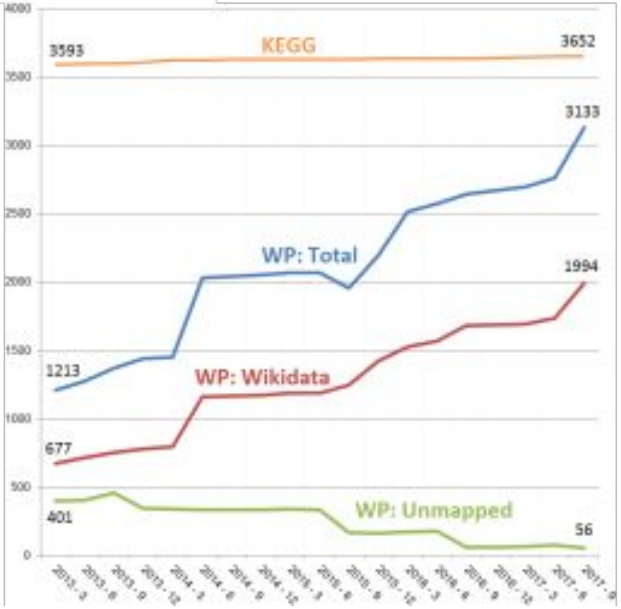
2015

source	count
HMDB	569
ChEBI	496
KEGG Compound	408
CAS	293
PubChem-compound	217
Chemspider	156
PubChem-substance	24
LIPID MAPS	11
Wikipedia	9
ChemIDplus	7
Reactome	4
ChEMBL compound	2
Other	1
CTD Chemical	1
ChemSpider	1

2012

source	count
HMDB	522
Kegg Compound	389
CAS	267
ChEBI	244
Entrez Gene	136
PubChem-compound	108
Chemspider	15
Wikipedia	11
PubChem-substance	8
ChemIDplus	7
ChEMBL compound	2
3DMET	1
LIPID MAPS	1

Curated subset



Maastricht University

Continues Integration with Jenkins

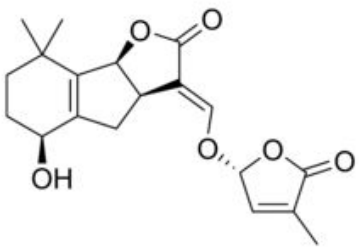
```
http://identifiers.org/chebi/CHEBI:36702 (1-alkyl-2-acyl-sn-glycero-3-phosphocholine (Plasmanylcholine)) does not have a Wikidata mapping
http://www.wikipathways.org/instance/WP2533_r107133 ;
http://identifiers.org/chebi/CHEBI:36702 (1-alkyl-sn-glycero-3-phosphocholine (Lyso PAF)) does not have a Wikidata mapping in
http://www.wikipathways.org/instance/WP2533_r107133 ;
http://identifiers.org/chebi/CHEBI:36707 (1-alkyl-2-acetyl-sn-glycero-3-phosphocholine (Platelet Activating Factor, PAF)) does not have a
http://www.wikipathways.org/instance/WP2533_r107133 ;
http://identifiers.org/chebi/CHEBI:36712 (1-alkyl-2-acyl-sn-glycero- 3-phosphoethanolamine) does not have a Wikidata mapping in
http://www.wikipathways.org/instance/WP2533_r107133 ;
http://identifiers.org/chebi/CHEBI:37296 (1-alkyl-2-acyl-sn-glycerol 3-phosphate (Plasmanic acid)) does not have a Wikidata mapping in
http://www.wikipathways.org/instance/WP2533_r107133 ;
http://identifiers.org/chebi/CHEBI:49172 (DAG) does not have a Wikidata mapping in http://www.wikipathways.org/instance/WP2889_r107138 ;
http://identifiers.org/chebi/CHEBI:49183 (Phosphatidylcholines) does not have a Wikidata mapping in http://www.wikipathways.org/instance/
http://identifiers.org/chebi/CHEBI:49183 (Phosphorylcholine) does not have a Wikidata mapping in http://www.wikipathways.org/instance/WP2
http://identifiers.org/chebi/CHEBI:58178 (Phosphatidylinositol-4-phosphate (PIP)) does not have a Wikidata mapping in http://www.wikipath
http://identifiers.org/chebi/CHEBI:60836 (PC) does not have a Wikidata mapping in http://www.wikipathways.org/instance/WP2579_r107043 ;
http://identifiers.org/chebi/CHEBI:63562 (GR ligand) does not have a Wikidata mapping in http://www.wikipathways.org/instance/WP2880_r875
http://identifiers.org/chebi/CHEBI:63562 (Ligand) does not have a Wikidata mapping in http://www.wikipathways.org/instance/WP2880_r87558
http://identifiers.org/chebi/CHEBI:68487 (Strigolactone) does not have a Wikidata mapping in http://www.wikipathways.org/instance/WP2945_
http://identifiers.org/chebi/CHEBI:76617 (JAK-STAT) does not have a Wikidata mapping in http://www.wikipathways.org/instance/WP2059_r1067
http://identifiers.org/chebi/CHEBI:77318 (Ligand) does not have a Wikidata mapping in http://www.wikipathways.org/instance/WP2876_r106535
http://identifiers.org/chebi/CHEBI:77326 (Ligand) does not have a Wikidata mapping in http://www.wikipathways.org/instance/WP2875_r106366
http://identifiers.org/chebi/CHEBI:78682 (fructose 1,6 Bisphosphate) does not have a Wikidata mapping in http://www.wikipathways.org/inst
http://identifiers.org/chebi/CHEBI:78682 (fructose 1,6 Bisphosphate) does not have a Wikidata mapping in http://www.wikipathways.org/inst
http://identifiers.org/chebi/CHEBI:78697 (fructose-6-phosphate) does not have a Wikidata mapping in http://www.wikipathways.org/instance/
http://identifiers.org/chebi/CHEBI:78697 (fructose-6-phosphate) does not have a Wikidata mapping in http://www.wikipathways.org/instance/
http://identifiers.org/chebi/CHEBI:78697 (fructose-6-phosphate) does not have a Wikidata mapping in http://www.wikipathways.org/instance/
http://identifiers.org/chebi/CHEBI:80219 (Dihydro- lipoamide-E) does not have a Wikidata mapping in http://www.wikipathways.org/instance/
http://identifiers.org/chebi/CHEBI:80233 (ANP) does not have a Wikidata mapping in http://www.wikipathways.org/i
http://identifiers.org/chebi/CHEBI:80234 (BNP) does not have a Wikidata mapping in http://www.wikipathways.org/i
http://identifiers.org/chebi/CHEBI:80235 (CNP) does not have a Wikidata mapping in http://www.wikipathways.org/i
http://identifiers.org/chebi/CHEBI:80337 (CGRP) does not have a Wikidata mapping in http://www.wikipathways.org/
http://identifiers.org/chebi/CHEBI:80339 (Adrenomedullin) does not have a Wikidata mapping in http://www.wikipat
http://identifiers.org/chebi/CHEBI:86029 (LXR ligand) does not have a Wikidata mapping in http://www.wikipathway
http://identifiers.org/chebi/CHEBI:86029 (Ligand) does not have a Wikidata mapping in http://www.wikipathways.or
--> expected: <0> but was: <45>
```

Maastricht University

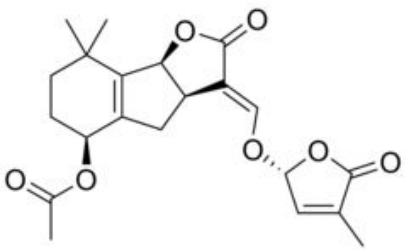
Strigolactones (in Wikipedia)

Chemical structures [edit]

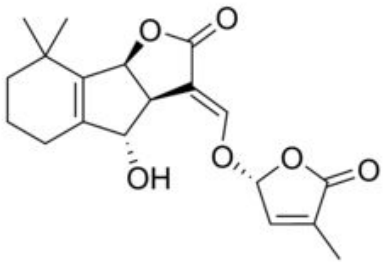
Some examples of strigolactones include:



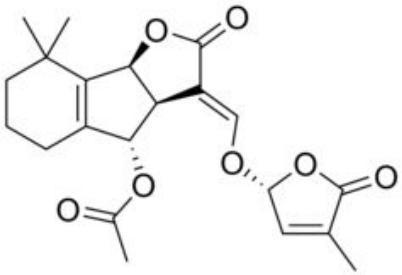
(+)-Strigol



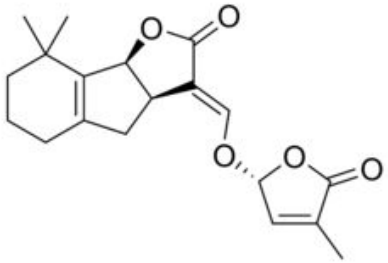
(+)-Strigyl acetate



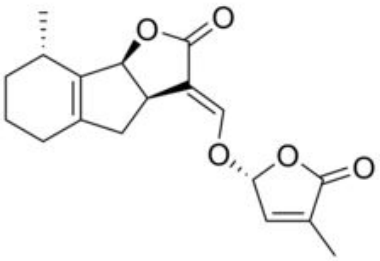
(+)-Orobanchol



(+)-Orobanchyl acetate



(+)-5-Deoxystrigol

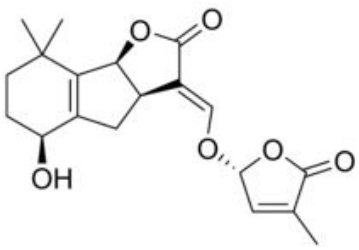


Sorgolactone

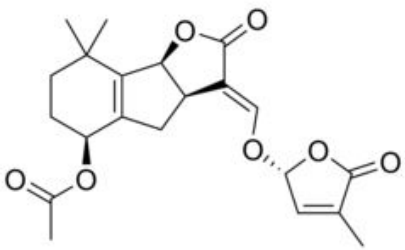
Strigolactones (in Wikipedia and Wikidata?)

Chemical structures [edit]

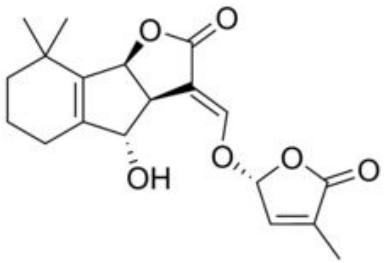
Some examples of strigolactones include:



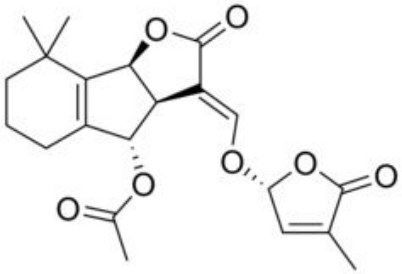
(+)-Strigol



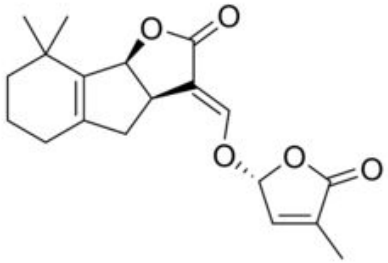
(+)-Strigyl acetate



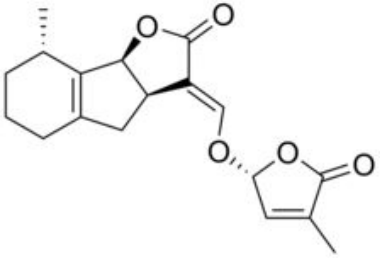
(+)-Orobanchol



(+)-Orobanchyl acetate



(+)-5-Deoxystrigol



Sorgolactone

strigolactones (Q2157332)

Strigolactones are a group of chemical compounds produced by a plant's roots. Due to their mechanism of actio plant hormones or phytohormones. So far, strigolactones have been identified to be responsible for three differer promote the germination of parasitic organisms that grow in the host plant's roots, such as Striga lutea and othe English Wikipedia)

Class Hierarchy

Wikidata / Scholia

Redirecting

If you know the identifier then Scholia can make a lookup based on the identifier:

cas/50-00-0

Lookup CAS 50-00-0. This will identify formaldehyde and redirect to its Scholia page.

inchikey/QTBSBXVTEAMEQO-UHFFFAOYSA-N

Redirect also works for InChIKeys, here for acetic acid.

Show 10 entries Search:

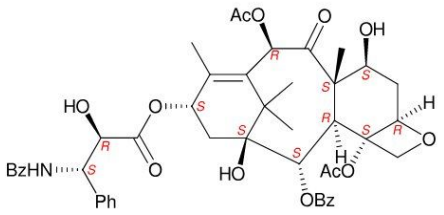
Mol	InChIKey	CAS	ChemSpider	PubChem CID
acetic acid	QTBSBXVTEAMEQO-UHFFFAOYSA-N	64-19-7	171	176
deuterated acetic acid	QTBSBXVTEAMEQO-GUEYOVJQSA-N	1186-52-3	2006083	2723903
acetic acid c-14	QTBSBXVTEAMEQO-HQMMCQRPSA-N	2845-03-6	144444	164769
acetic acid c-13	QTBSBXVTEAMEQO-VQEHIDDOSA-N	1563-79-7	8329490	10153982
acetic acid c-11	QTBSBXVTEAMEQO-JVVVGQRLSA-N	78887-71-5	396653	450349
acetate ion	QTBSBXVTEAMEQO-UHFFFAOYSA-M	71-50-1	170	175

Edit on query.Wikidata.org

Showing 1 to 6 of 6 entries Previous 1 Next

paclitaxel (Q423762)

Paclitaxel (PTX), sold under the brand name Taxol among others, is a chemotherapy medication used to treat a number of types of cancer. This includes ovarian cancer, breast cancer, lung cancer, Kaposi sarcoma, cervical cancer, and pancreatic cancer. It is given by injection into a vein. ... (from the English Wikipedia)



2019: 10.3897/rio.5.e35820
2017: 10.6084/m9.figshare.6356027.v1

Maastricht University

Identifiers

Show 10 entries Search:

IDpred	Id

SCHOLIA

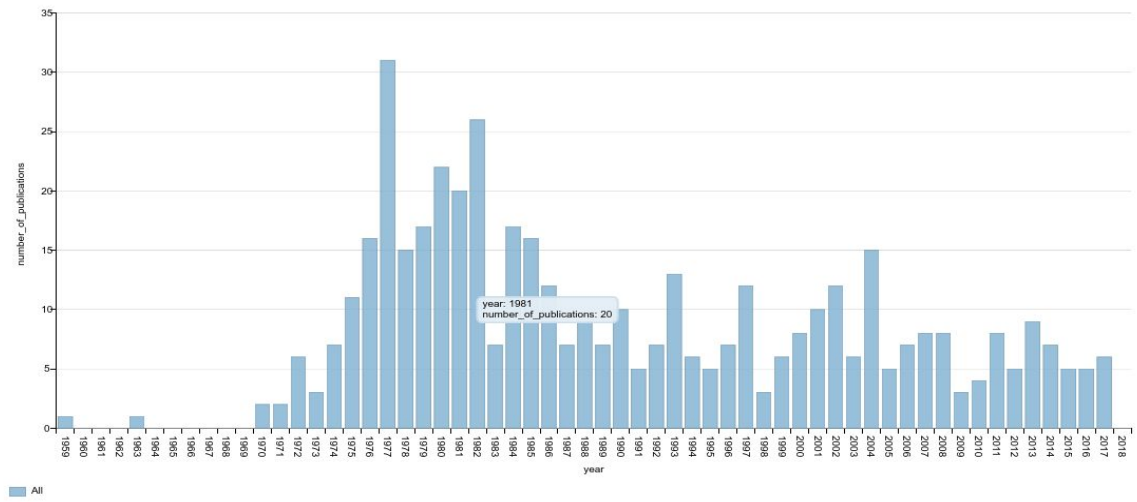
Wikidata / Scholia

Physchem Properties

Show 10 entries Search:

PropEntity	Value	Units	Qualifiers	Source	Doi
acid dissociation constant	4.74	1		Small Scale Determination of the pKa Values for Organic Acids	10.1021/ED071PA6
mass	60.021129	atomic mass unit		PubChem	
acid dissociation constant	4.756	1	temperature: 25	CRC Handbook of Chemistry and Physics (95th edition)	
boiling point	117.9	degrees Celsius	pressure: 101325	CRC Handbook of Chemistry and Physics (95th edition)	
density	1.0446	gram per cubic centimetre	temperature: 25	CRC Handbook of Chemistry and Physics (95th edition)	

Publications per year



Recently published works on the chemical

Show 10 entries

Date	Work	Type	Topics
2017-08-09	In vitro human skin permeation of benzene in gasoline: effects of concentration, multiple dosing and skin preparation	scholarly article	oil and gas extraction // benzene
2017-04-27	Nicotine, aerosol particles, carbonyls and volatile organic compounds in tobacco- and menthol-flavored e-cigarettes	scholarly article	toluene // benzene

Examples: OECD Sections and tests

Scholia Author Work Organization Location Event Project Award Topic Tools Help

venue

OECD Guidelines for the Testing of Chemicals, Section 1 (Q57978040)

Recently published works

Show 10 entries Search:

Publication date	Work	Authors
2012-10-02	Test No. 109: Density of Liquids and Solids	
2000-01-21	Test No. 106: Adsorption -- Desorption Using a Batch Equilibrium Method	
1995-07-27	Test No. 102: Melting Point/ Melting Range	
1995-07-27	Test No. 105: Water Solubility	
1981-05-12	Test No. 113: Screening Test for Thermal Stability and Stability in Air	
1981-05-12	Test No. 116: Fat Solubility of Solid and Liquid Substances	

Edit on query.Wikidata.org

Showing 1 to 6 of 6 entries Previous 1 Next

Topics

Show 10 entries Search:

Count	Topic	Example work
2	solid	Test No. 109: Density of Liquids and Solids
2	liquid	Test No. 109: Density of Liquids and Solids

Scholia Author Work Organization Location Event Project Award Topic Tools Help

Test No. 109: Density of Liquids and Solids (Q60233153)

Show 10 entries Search:

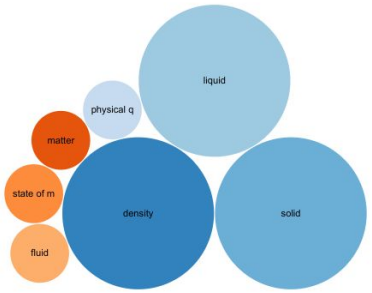
Order	Author	Orcid
No data available in table		

Edit on query.Wikidata.org

Showing 0 to 0 of 0 entries Previous Next

Topic scores

Topics based on a weighting between main subject of work, cited and citing works.



NanoCommons Nano-Knowledge Community

Maastricht University

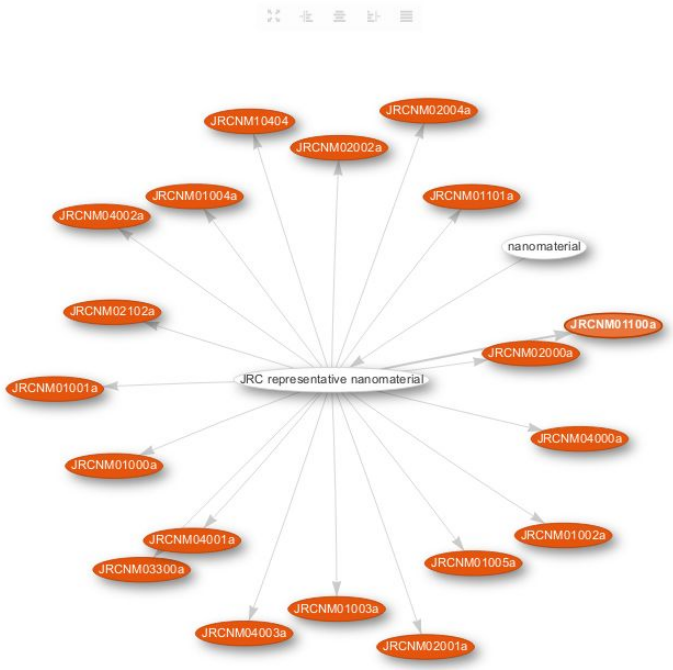
SCHOLIA

Scholia: JRC representative industrial nanomaterials

topic | chemical

JRC representative nanomaterial (Q47461491)

Class Hierarchy



Recently published works on the chemical RSS

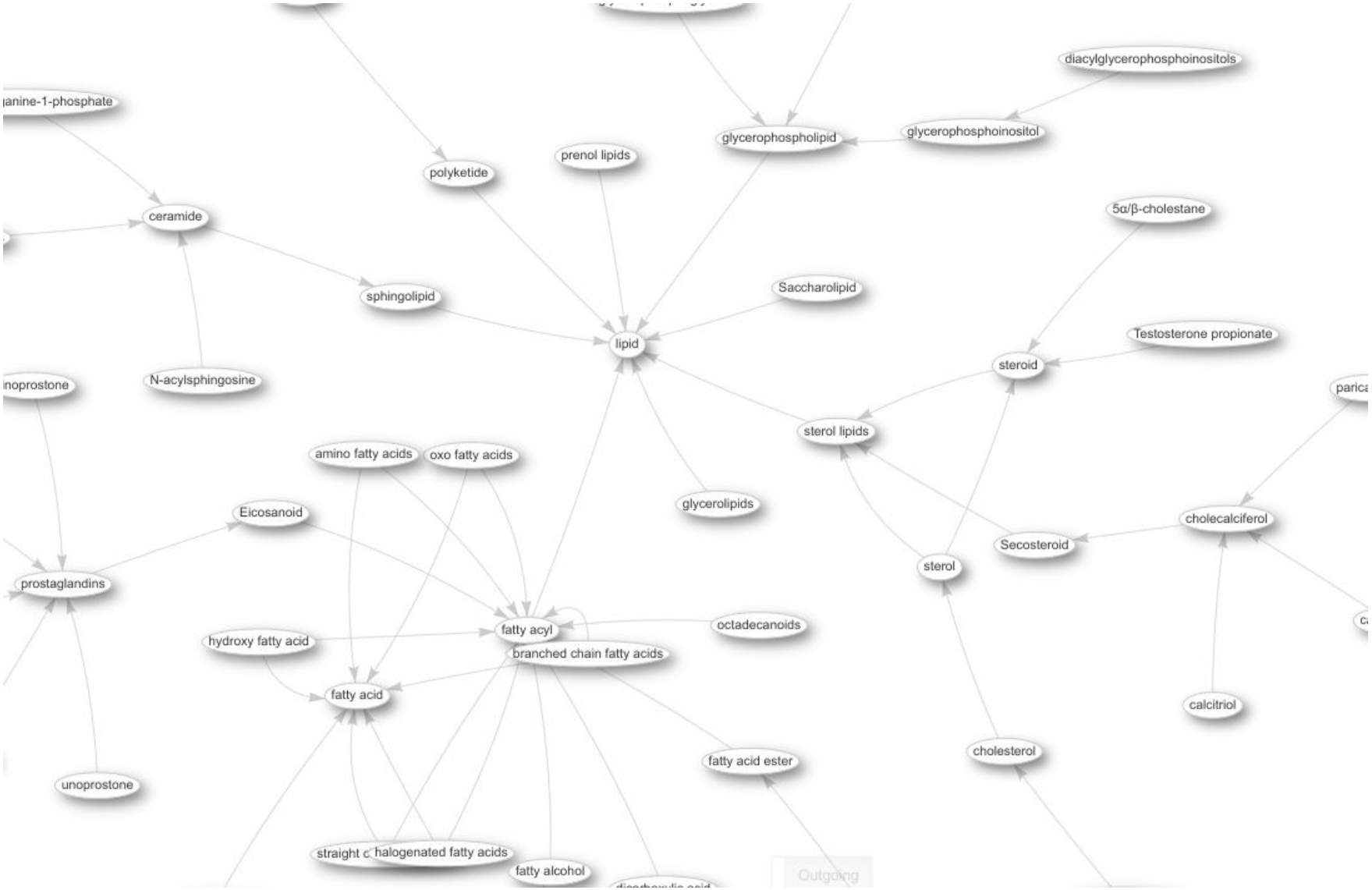
Show 10 entries Search:

Date	Work	Type	Topics
2017-09-28	Fish cell lines as a tool for the ecotoxicity assessment and ranking of engineered nanomaterials.	scholarly article	JRCNM02000a // JRCNM04000a // JRCNM01101a // JRCNM01100a // JRCNM02102a // nanomaterial // toxicology
2017-06-01	Graphistrength© C100 MultiWalled Carbon Nanotubes (MWCNT): thirteen-week inhalation toxicity study in rats with 13- and 52-week recovery periods combined with comet and micronucleus assays	scholarly article	JRCNM04002a // Brown Rat // toxicology
2017-05-19	Elucidating the Role of Dissolution in CeO2 Nanoparticle Plant Uptake by Smart Radiolabeling.	scholarly article	JRCNM02102a // general chemistry // catalysis // nanoparticle
2017-04-05	Multi-walled carbon nanotube-physicochemical properties predict the systemic acute phase response following pulmonary exposure in mice.	scholarly article	JRCNM04003a // JRCNM04001a // JRCNM04000a // carbon nanotube
2017-01-03	Negligible cytotoxicity induced by different titanium dioxide nanoparticles in fish cell lines.	scholarly article	JRCNM01005a // JRCNM01004a // JRCNM01003a
2016-11-01	The JRC Nanomaterials Repository: A unique facility providing representative test materials for nanoEHS research	scholarly article	JRC representative nanomaterial // Directorate-General for Joint Research Centre // nanomaterial // toxicology
2015-11-12	Towards the standardization of nanoecotoxicity testing: Natural organic matter 'camouflages' the adverse effects of TiO2 and CeO2 nanoparticles on green microalgae.	scholarly article	JRCNM02102a // JRCNM01003a

SCHOLIA

{ } wikicite

The LIPID MAPS hierarchy (in Wikidata)



Maastricht University

w.wiki/3a3

class	classLabel	lmid	count
wd:Q63433687	fatty acyl	LMFA	0
wd:Q63434442	straight chain fatty acids	LMFA0101	37
wd:Q24901874	branched chain fatty acids	LMFA0102	79
wd:Q61737535	unsaturated fatty acid	LMFA0103	279
wd:Q40211102	hydroxy fatty acid	LMFA0105	184
wd:Q63435564	oxo fatty acids	LMFA0106	56
wd:Q63436532	halogenated fatty acids	LMFA0109	24
wd:Q63434663	amino fatty acids	LMFA0110	39
wd:Q422050	dicarboxylic acid	LMFA0117	78
wd:Q61716319	octadecanoids	LMFA02	82
wd:Q407680	Eicosanoid	LMFA03	83
wd:Q209717	prostaglandins	LMFA0301	89
wd:Q4198767	isoprostane	LMFA0311	5
wd:Q378871	fatty alcohol	LMFA05	156

In which species is this lipid found?

lipid	lipidLabel	lmid	species	speciesLabel	source	sourceLabel	doi
wd:Q26840883	(-)-methyl jasmonate	LMFA02020010	wd:Q23501	Solanum lycopersicum	wd:Q33228063	Induced defences in plants reduce herbivory by increasing cannibalism	10.1038/S41559-017-0231-6
wd:Q27158341	quercetin 5,7,3',4'-tetramethyl ether	LMPK12112771	wd:Q22701	Sambucus nigra	wd:Q39812430	Elderberry flavonoids bind to and prevent H1N1 infection in vitro.	10.1016/J.PHYTOCHEM.2009.06.003
wd:Q55620521	(R)-1,7-Dioxaspiro[5.5]undecane	LMPK09000012	wd:Q2207329	olive fruit fly	wd:Q55645881	Sex-specific activity of (R)-(-)- and (S)-(+)-1,7-dioxaspiro[5.5]undecane, the major pheromone of Dacus oleae	10.1007/BF01012372
wd:Q55620476	(S)-1,7-Dioxaspiro[5.5]undecane	LMPK09000013	wd:Q2207329	olive fruit fly	wd:Q55645881	Sex-specific activity of (R)-(-)- and (S)-(+)-1,7-dioxaspiro[5.5]undecane, the major pheromone of Dacus oleae	10.1007/BF01012372
wd:Q27135687	geranylacetone	LMFA11000696	wd:Q16528	Nelumbo nucifera	wd:Q902623	ChEBI	
wd:Q27135687	geranylacetone	LMFA11000696	wd:Q16528	Nelumbo nucifera	wd:Q43240571	Comparative analysis of essential oil components and antioxidant activity of extracts of Nelumbo nucifera from various	10.1021/JF902643E

Visualize Wikidata Schema

racemic mixture

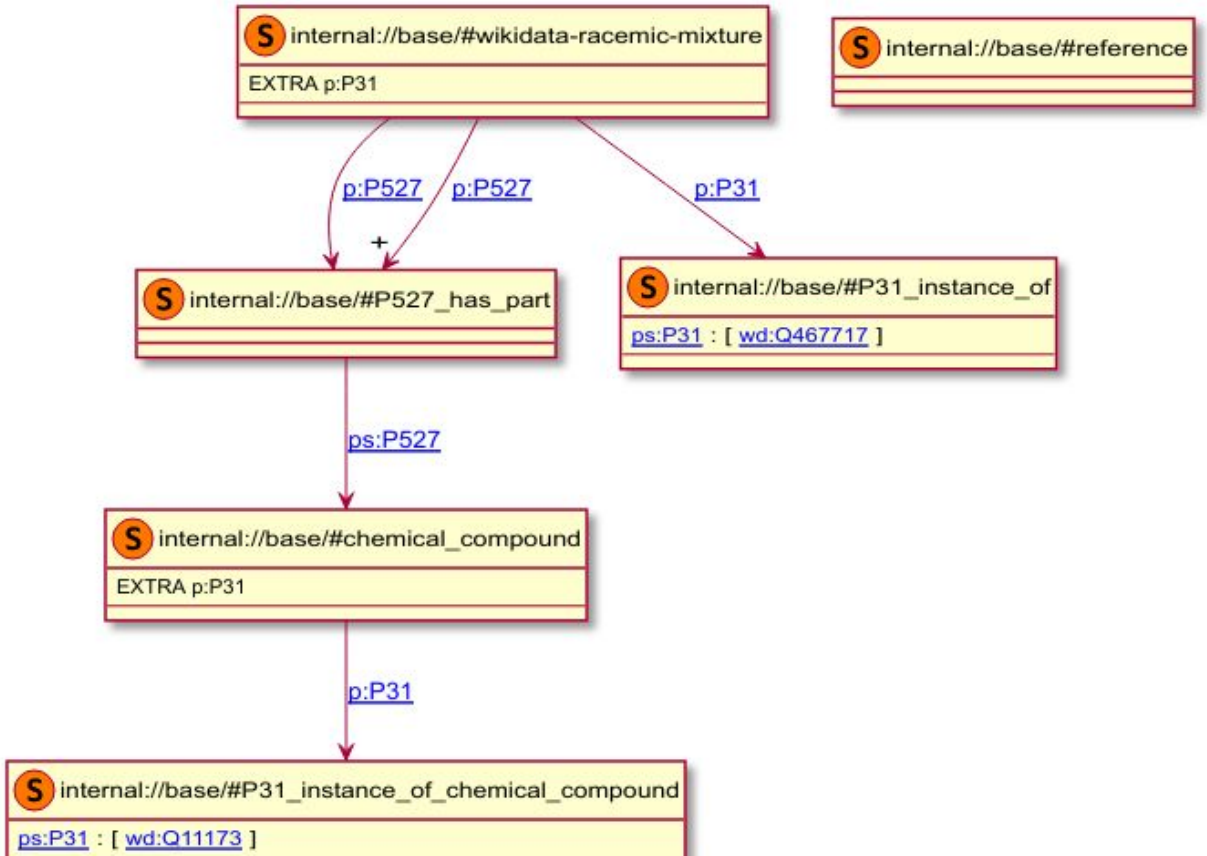
Language en

Info about schema entity

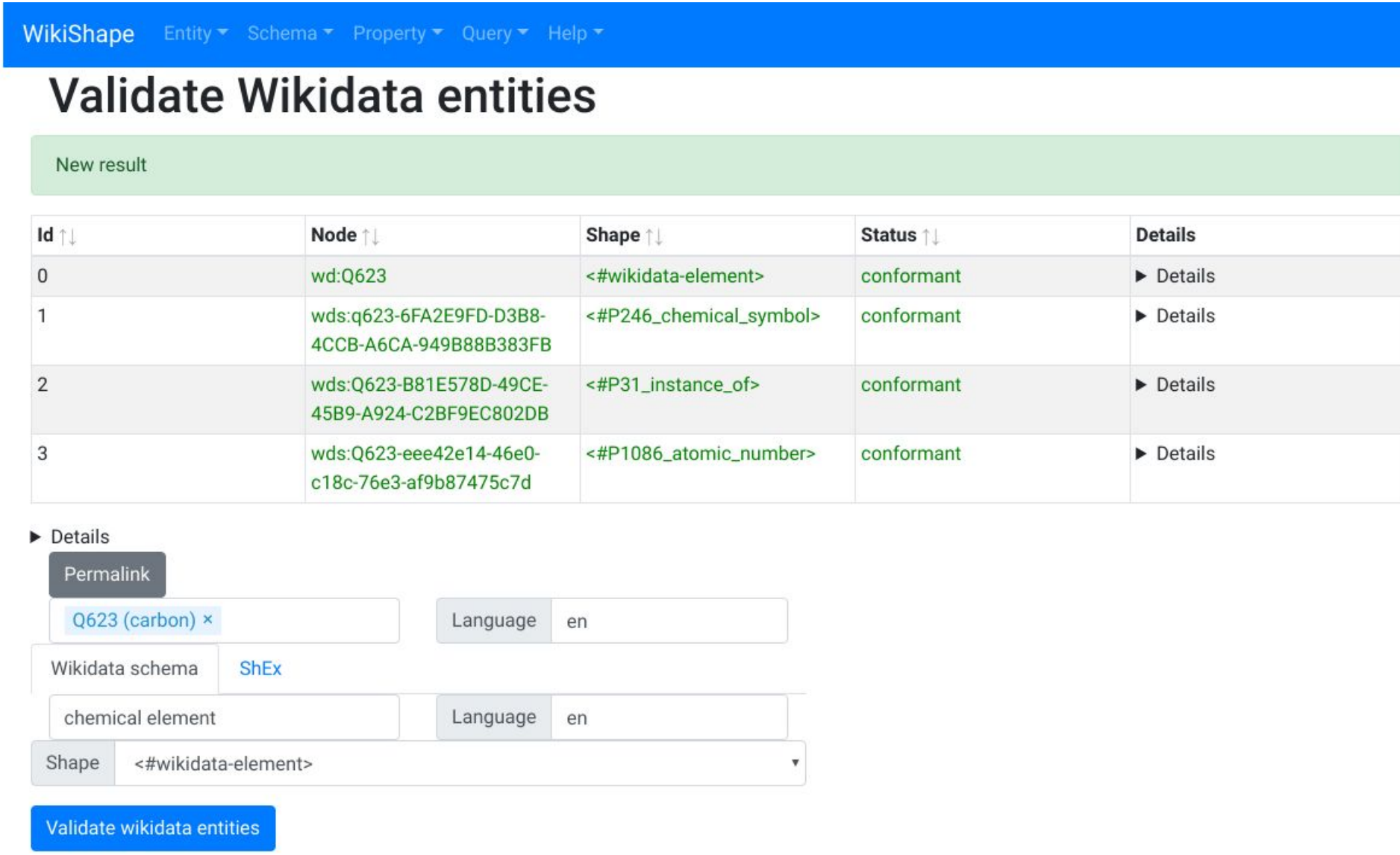
E47 - racemic mixture

mixture of chemicals with the same structure but different stereochemistry

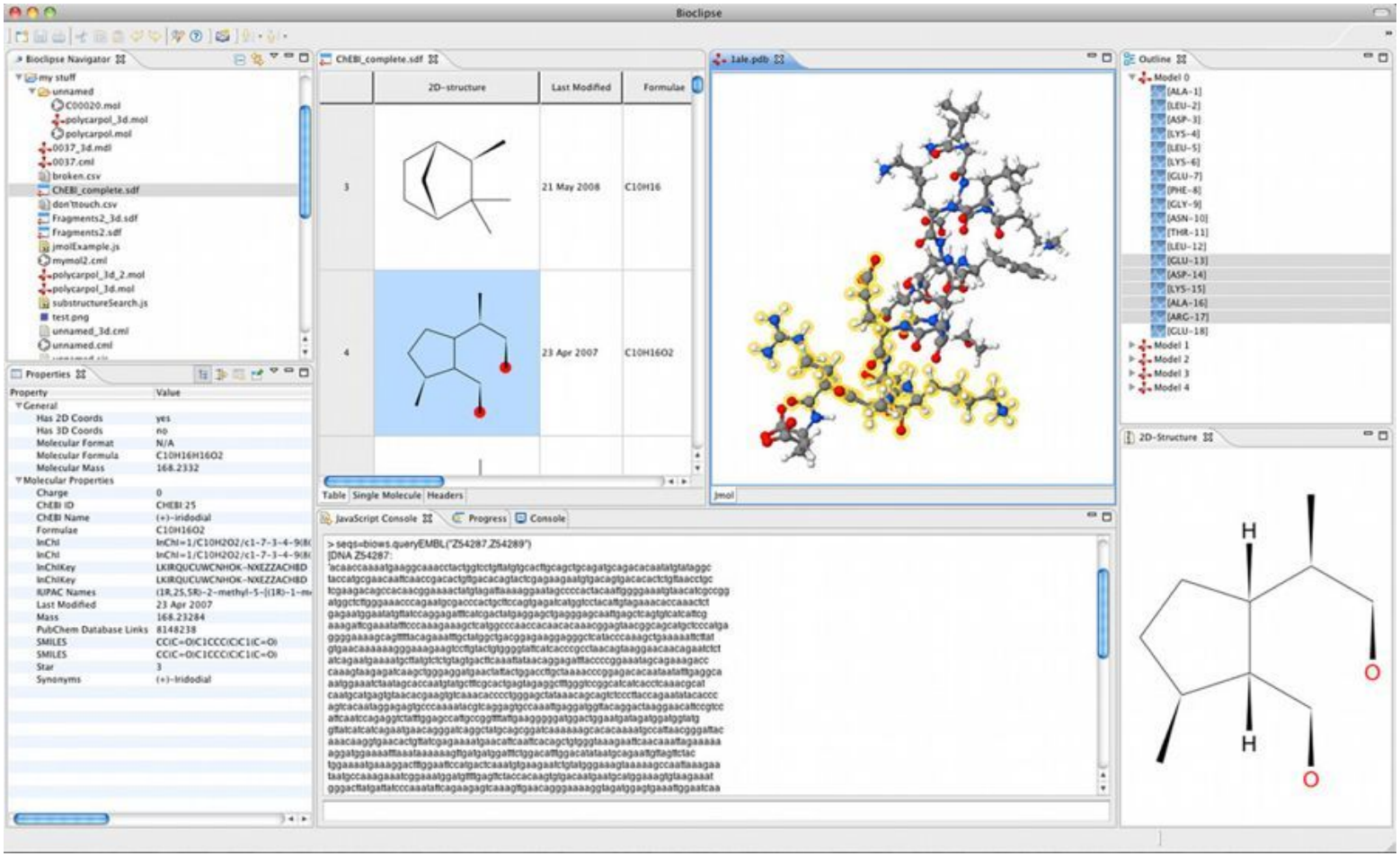
https://www.wikidata.org/wiki/EntitySchema:E47



ShEx validation: E46 → chemical element



Workhorse: Bioclipse scripts



10.1186/1471-2105-8-59,
10.1186/1471-2105-10-397

Maastricht University

Bacting: Bioclipse on the command line

```
@Grab(group='io.github.egonw.bacting', module='managers-cdk', version='0.0.9')

workspaceRoot = "."
def cdk = new net.bioclipse.managers.CDKManager(workspaceRoot);

println cdk.fromSMILES("COC")
```

- Wikicite/findConcepts.groovy
- Wikidata/createWDitemsFromSMILES.groovy
- LipidMaps/classifyLipids.groovy
- ExtIdentifiers/comptox.groovy
- MeltingPoints/createQuickStatements.groovy
- ...

https://github.com/egonw/bacting
https://github.com/egonw/ons-wikidata

Journal of Cheminformatics

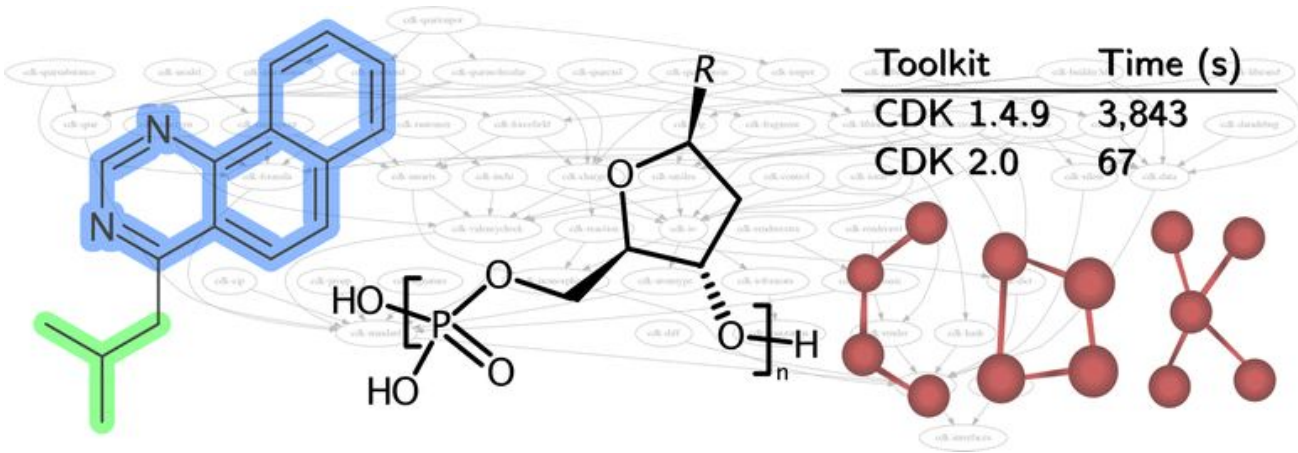
Software | Open Access | Published: 06 June 2017

The Chemistry Development Kit (CDK) v2.0: atom typing, depiction, molecular formulas, and substructure searching

Egon L. Willighagen, John W. Mayfield, Jonathan Alvarsson, Arvid Berg, Lars Carlsson, Nina Jeliazkova, Stefan Kuhn, Tomáš Pluskal, Miquel Rojas-Chertó, Ola Spjuth, Gilleain Torrance, Chris T. Evelo, Rajarshi Guha & Christoph Steinbeck

Journal of Cheminformatics **9**, Article number: 33 (2017) | Download Citation

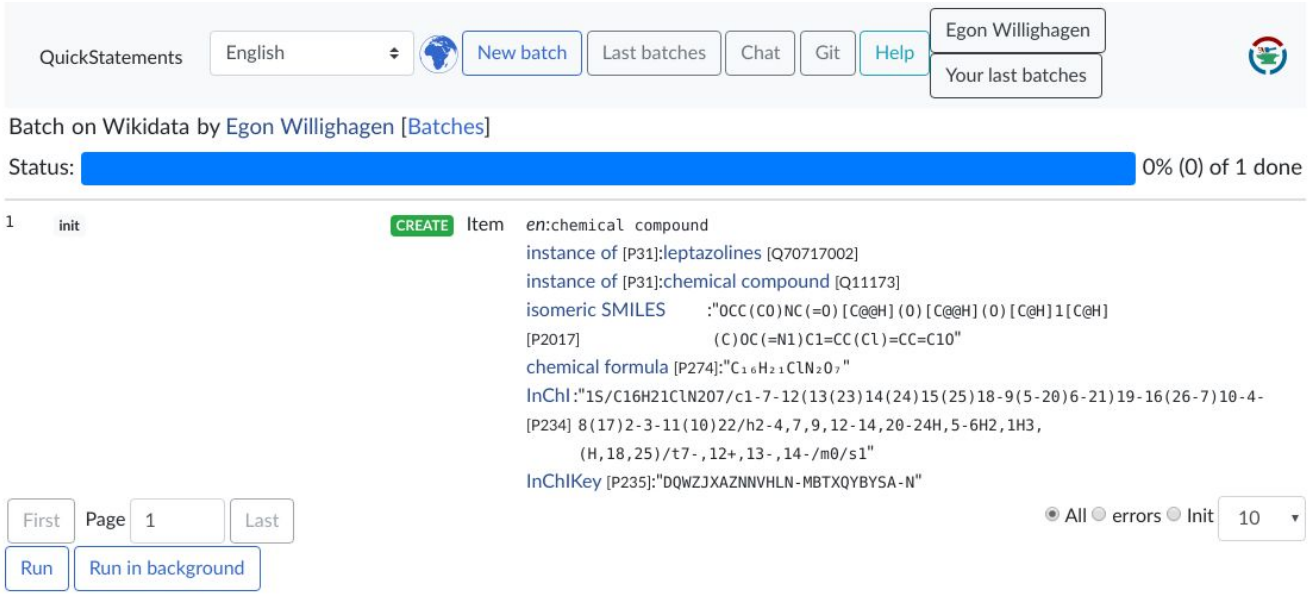
7825 Accesses | **50** Citations | **55** Altmetric | Metrics »



Wikidata Quickstatements v1

CREATE

LAST P31 Q70717002
LAST P31 Q11173
LAST Den "chemical compound"
LAST P2017 "OCC(CO)NC(=O)[C@@H](O)[C@@H](O)[C@H]1[C@..."
LAST P274 "$C_{16}H_{22}N_2O_7$"
LAST P234 "1S/C16H22N2O7/c1-8-12(13(22)14(23)15(24)17-9(6-..."
LAST P235 "XXWREFOKTUTPHF-ISTUKMMPSA-N"

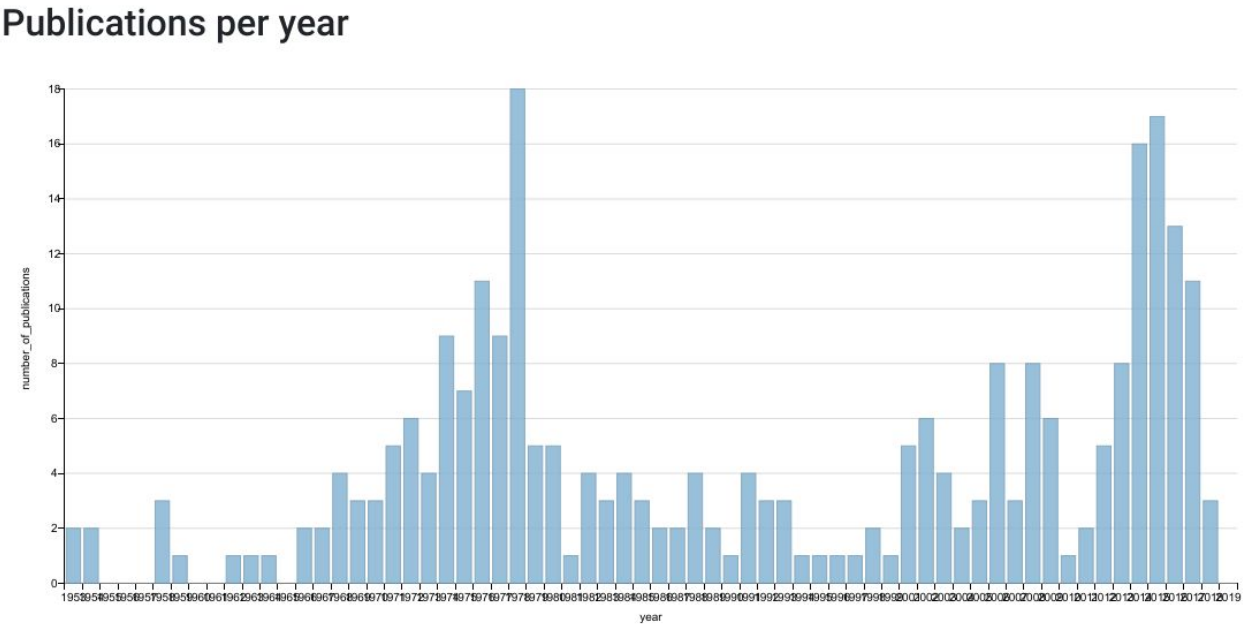


Dr. Magnus Manske
Sanger Institute

Wikidata Quickstatements v2

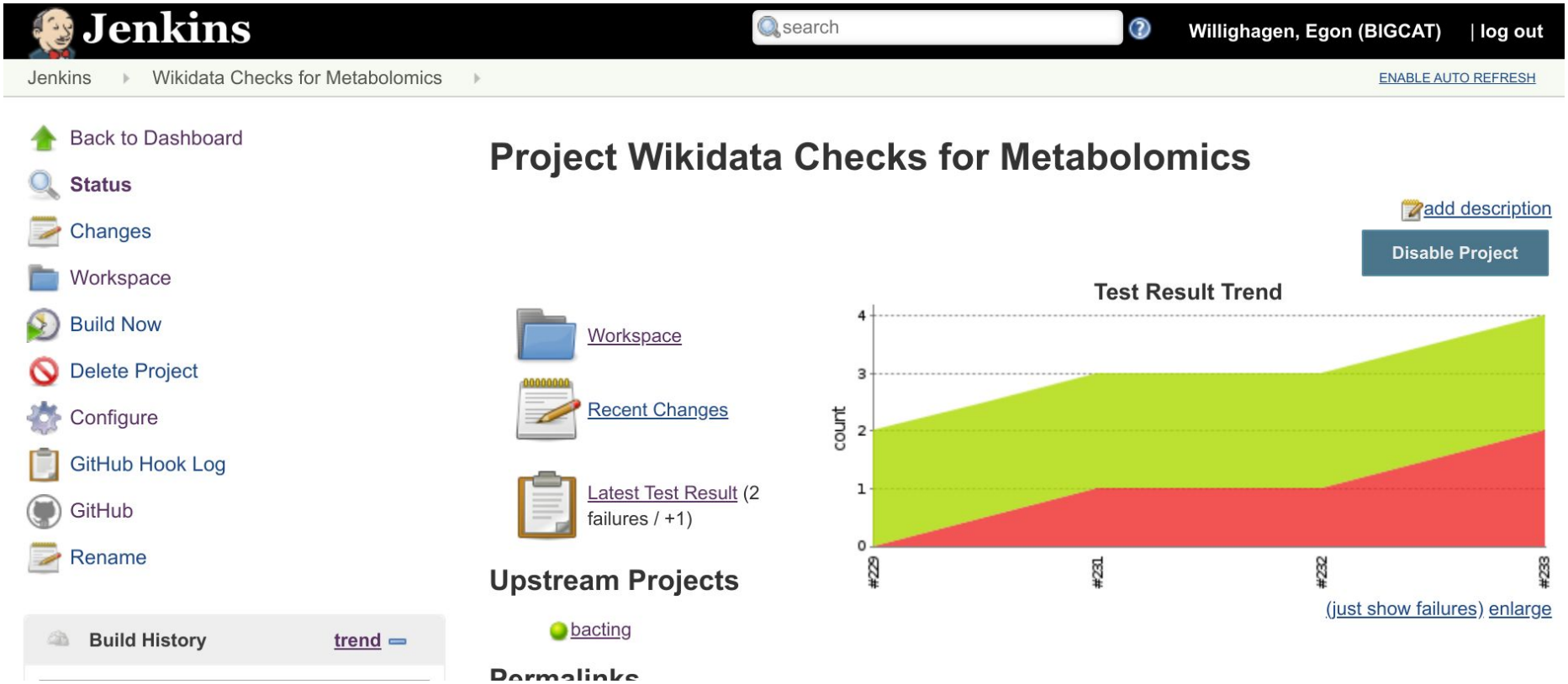
qid,P921,#
Q26801490,Q70828631,Activities and Effects of Ergot Alkaloids on ...
Q28082319,Q70828631,Diversification of ergot alkaloids in natural and ...
Q28214648,Q70828631,Biotechnology and genetics of ergot alkaloids
Q28276288,Q70828631,Ergot alkaloids--biology and molecular biology
Q28287164,Q70828631,Occurrence of peptide and clavine ergot alkaloids …
...

Dr. Magnus Manske
Sanger Institute



Maastricht University

Jenkins for Wikidata quality control



Maastricht University

ChemCuration example: InChIKeys

Jenkins › Wikidata Checks for Metabolomics › #233 › Test Results › (root) › InChITests › InChIKeyMismatch ENABLE AUTO REFRESH

History
Git Build Data
No Tags
Test Result
Previous Build

Error Message

The InChIKey computed from the isomeric SMILES and InChIKey in Wikidata does not match

Stacktrace

http://www.wikidata.org/entity/Q421291 with isomeric SMILES '[Fe+2].O[C@H]([C@@H](O)C([O-])=O)[C@H](O)[C@H](O)CO.[O-]C(=O)[C@H](O)[C@@H](O)[C@H](O)[C@H](O)CO' has a calculated InChIKey VRIVJOXICYMTAG-IYEMJOQQSA-L that does not match the given QDUZQOIJXPPTLY-GMBKLUGCSA-N
http://www.wikidata.org/entity/Q7777226 with isomeric SMILES 'Oc1cc(cc(O)c1O)C(=O)Oc5c(O)cc([C@H]2Oc3cc(O)cc(O)c3C[C@H]2O)c6\C=C(/C=C(/OC(=O)c4cc(O)c(O)c(O)c4)C(=O)c56)[C@H]7Oc8cc(O)cc(O)c8C[C@H]7O' has a calculated InChIKey FJYGFTHLNNSVPY-BBXLVSEPSA-N that does not match the given TUJOKWPTOVJHLY-JBJHRQGLSA-N
http://www.wikidata.org/entity/Q15427926 with isomeric SMILES 'CC1(C)C([C@@H](OC(C)=O)C[C@@]2(C)[C@](C([C@@H](OC(C)=O)CC2)=C)([H])[C@@H]3OC(C)=O)=C(C)[C@@H](OC([C@H](C)[C@H](C)O)=O)C[C@]13[H]' has a calculated InChIKey ULHOQEOTAJVICR-SJJKDWJASA-N that does not match the given FMPIEMVVEJGMCY-IRWPHOLZSA-N
http://www.wikidata.org/entity/Q568 with isomeric SMILES '[Li]' has a calculated InChIKey WHXSMMKQMYFTQS-UHFFFAOYSA-N that does not match the given SIAPCJWMELPYOE-UHFFFAOYSA-N
http://www.wikidata.org/entity/Q5278705 with isomeric SMILES 'C[C@@]13O[C@]1(/C=C/C(C)=C/C=C/C(C)=C/C=C/C=C(C)/C=C/C=C(C)/[C@@H]=C=C2C(C)(C)C[C@@H](OC(C)=O)C[C@]2(C)O)C(C)(C)C[C@H](O)C3' has a calculated InChIKey PVNVIBOWBAPFOE-RWNIHPGNSA-N that does not match the given GJFBHWJTMDTLNX-UWCSZFODSA-N

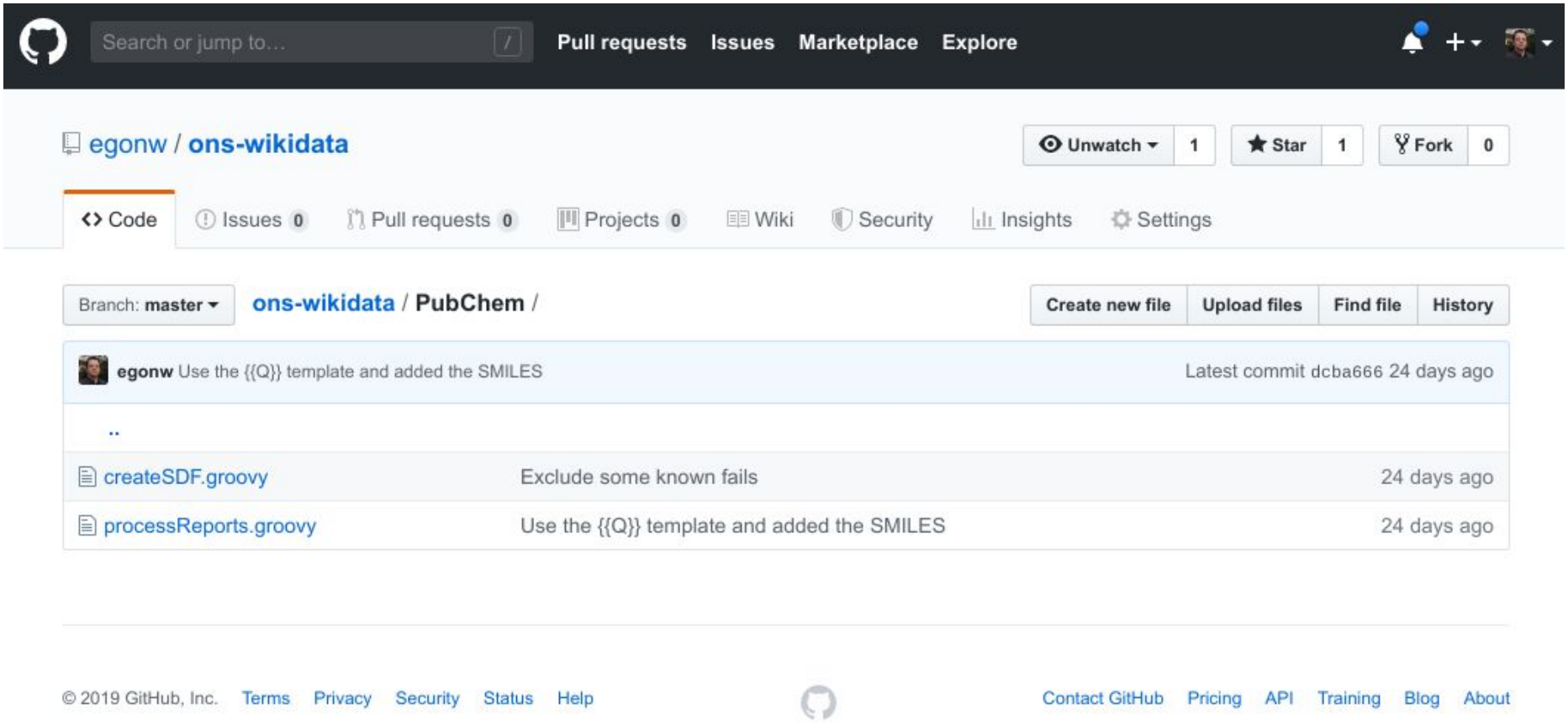
Wikidata in PubChem (ongoing)

PubChem deposit [edit | Add topic]

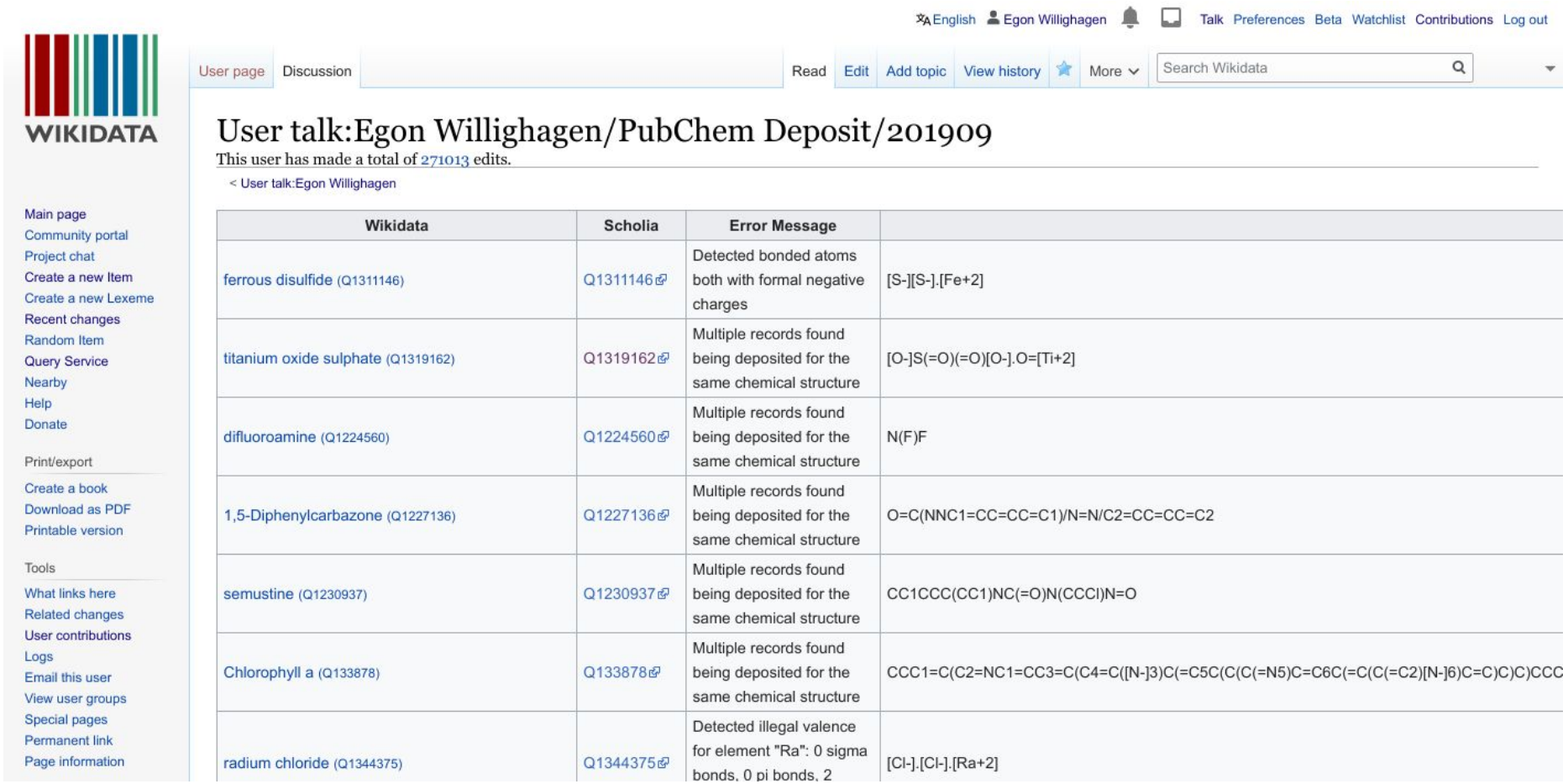
Notified participants of WikiProject Chemistry Hi all, I want to let everyone know that I have initiated uploading the chemicals from Wikidata to PubChem. This will create a further route to crosslink the databases (Wikidata and Wikipedia already link to PubChem, Wikipedia is actively being deposited in PubChem). Now, Wikipedia != Wikidata and uploading Wikidata separately actually has additional advantages, such as further validation reports. I already fixed a number of SMILES errors found by PubChem and not by the Chemistry Development Kit. It also reports duplicated, and a lot more. I will upload the report somewhere as soon as I have it. I have created a script to create an input CSV file (https://github.com/egonw/ons-wikidata/blob/master/PubChem/createSDF.groovy). More later. --Egon Willighagen (talk) 16:18, 22 September 2019 (UTC)

> Update: the first deposit is committed and now up for review with PubChem curators. I got two reports, but neither contain the external identifier, so I need to combine these with the input first before they are useful. More later. --Egon Willighagen (talk) 17:22, 22 September 2019 (UTC)
>
> > Update: and here are the reports (created with https://github.com/egonw/ons-wikidata/blob/master/PubChem/processReports.groovy): https://www.wikidata.org/wiki/User_talk:Egon_Willighagen/PubChem_Deposit/201909 --Egon Willighagen (talk) 18:41, 22 September 2019 (UTC)
> >
> > > I am having trouble following. I think you are saying that currently Wikidata items and PubChem items map to each other on the wiki side, but not on the PubChem side, and you are sharing information on the PubChem side so that people can start there and navigate to wiki. If this is correct, then that seems great.
> > >
> > > Currently you are treating Wikidata and Wikipedia as different entities because even though Wikidata and Wikipedia link to each other, their content is different enough to justify two links. Also, the PubChem community is unlikely to know how to readily move from one to the other, so that is another reason for two links. You shared your mapping software in GitHub. You have a log of error reports published in a table on wiki.
> > >
> > > This all seems useful, so great. **Blue Rasberry** (talk) 15:26, 23 September 2019 (UTC)
> > >
> > > > @Egon Willighagen: If you have good contact with PubChem, could you asked them to generate a subset of their data containing PubChem CID, InChI, InChKey and SMILES under CC0 ? MAin argument: if all databases are doing the same, WD can becomes the way for databases to access to chemical IDs in other databases.
> > > >
> > > > Currently only DrugBank played the game. Snipre (talk) 11:52, 27 September 2019 (UTC)
> > > >
> > > > > Yes, will ask Evan soon. We'll both be at the Beilstein Open Science meeting. In the past the answer was: PubChem is public domain and cannot have a CC0 license/waiver (which claims ownership). The other problem is to determine which parts of PubChem are public domain, and which are owned by the data provider :(--Egon Willighagen (talk) 17:55, 27 September 2019 (UTC)

Wikidata in PubChem (ongoing)



Wikidata in PubChem (ongoing)



English Egon Willighagen Talk Preferences Beta Watchlist Contributions Log out

WIKIDATA

User page Discussion Read Edit Add topic View history More Search Wikidata

User talk:Egon Willighagen/PubChem Deposit/201909

This user has made a total of 271013 edits.

< User talk:Egon Willighagen

Main page
Community portal
Project chat
Create a new Item
Create a new Lexeme
Recent changes
Random Item
Query Service
Nearby
Help
Donate

Print/export
Create a book
Download as PDF
Printable version

Tools
What links here
Related changes
User contributions
Logs
Email this user
View user groups
Special pages
Permanent link
Page information

Wikidata	Scholia	Error Message	
ferrous disulfide (Q1311146)	Q1311146	Detected bonded atoms both with formal negative charges	[S-][S-].[Fe+2]
titanium oxide sulphate (Q1319162)	Q1319162	Multiple records found being deposited for the same chemical structure	[O-]S(=O)(=O)[O-].O=[Ti+2]
difluoroamine (Q1224560)	Q1224560	Multiple records found being deposited for the same chemical structure	N(F)F
1,5-Diphenylcarbazone (Q1227136)	Q1227136	Multiple records found being deposited for the same chemical structure	O=C(NNC1=CC=CC=C1)/N=N/C2=CC=CC=C2
semustine (Q1230937)	Q1230937	Multiple records found being deposited for the same chemical structure	CC1CCC(CC1)NC(=O)N(CCCl)N=O
Chlorophyll a (Q133878)	Q133878	Multiple records found being deposited for the same chemical structure	CCC1=C(C2=NC1=CC3=C(C4=C([N-]3)C(=C5C(C(C(=N5)C=C6C(=C(C(=C2)[N-]6)C=C)C)C)CCC
radium chloride (Q1344375)	Q1344375	Detected illegal valence for element "Ra": 0 sigma bonds, 0 pi bonds, 2	[Cl-].[Cl-].[Ra+2]

Poster

figshare Browse Search on figshare... Log in Sign up

Maastricht University

NUTRIM

NUTRIM School of Nutrition and Translational Research in Metabolism

Wikidata and Scholia as a hub linking chemical knowledge

Egon Willighagen[A], Denise Slenter[A], Daniel Mietchen[B], Chris Evelo[A,C], Finn Nielsen[D]

[A] Department of Bioinformatics - BiGCaT, Maastricht University, The Netherlands, [B] Data Science Institute, University of Virginia, Charlottesville, Virginia, USA,
[C] Maastricht Centre for Systems Biology - MaCSBio, Maastricht University, The Netherlands, [D] Cognitive Systems, DTU Compute, Technical University of Denmark, Denmark

Introduction

Making chemical databases more FAIR (findable, accessible, interoperable, and reusable) benefits computational chemistry and cheminformatics. We here discuss Wikidata, a young sister project of Wikipedia, with one key difference: it is a machine readable database, making it far more useful for interoperability of molecular databases in systems biology [1,2]. Thanks to the WikiProject Chemistry community on Wikidata, there is a growing amount of information about chemical compounds.

Methods

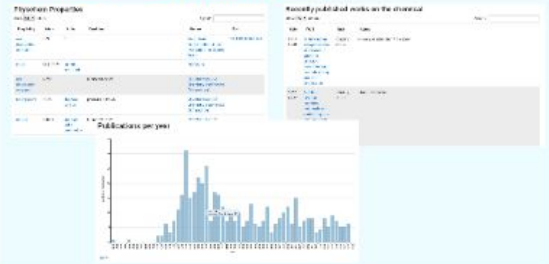
Scholia is a Python/Flask-based server system that creates webpages using a template approach [5]. It defines templates for concepts around knowledge exchange, such as publications, journals, publishers, but also topics. It uses SPARQL queries against the Wikidata Query Service (WDQS,

Results

We here introduce our contributions to the WikiProject Chemistry to support FAIR-ification of open chemical knowledge. For example, we proposed new Wikidata properties to annotate compounds with external database identifiers for the EPA CompTox Dashboard [3], the SPLASH [4], and MetaboLights. We also introduced a Scholia extension [5], visualizing data about chemicals and chemical classes:

https://tools.wmflabs.org/scholia/

Literature-backed (PhysChem) Facts



Linking Databases

Maastricht University

https://doi.org/10.6084/m9.figshare.6356027.v1

Acknowledgments

This work received funding from the European Union's Horizon 2020 research and innovation programme via the NanoCommons project under grant agreement No 731032 and eNanoMapper project under grant agreement No 604134, and from the Alfred P. Sloan Foundation under grant number G-2019-11458.

https://tools.wmflabs.org/scholia/