

Need and benefits for structure standardization to facilitate integration and connectivity between government databases

Valery Tkachenko², Christopher Grulke¹, Antony Williams¹

The views expressed in this presentation are those of the author and do not necessarily reflect the views or policies of the U.S. EPA

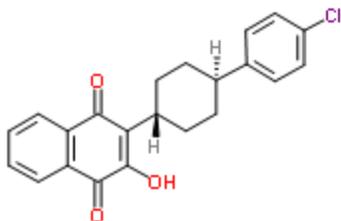
¹ National Center for Computational Toxicology, EPA, NC, United States;

² SCIENCE DATA SOFTWARE, Rockville, MD, United States

Fall ACS 2017, Washington, DC

ChemSpider – an example of chemical database

Search term: **atovaquone** (Found by approved synonym) ?



? 2D 3D Save Edit Zoom

 - 2 of 2 defined stereocentres

Atovaquone

ChemSpider ID: **10482034**

Molecular Formula: $C_{22}H_{19}ClO_3$

Average mass: 366.837494 Da

Monoisotopic mass: 366.102264 Da

▼ Systematic name

2-[trans-4-(4-Chlorophenyl)cyclohexyl]-3-hy

▶ SMILES and InChIs

▶ Cite this record

Wikibox

Embed

Deprecate

Watch this record

Manage data slice

▼ Names and Identifiers

Names and Synonyms Database ID(s)

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

(-)-Cholesterol

(3b)-cholest-5-en-3-ol

(3S,8S,9S,10R,13R,14S,17R)-10,13-Dimethyl-17-[(2R)-6-methyl-2-heptanyl]-2,3,4,7,8,9,10,11,12,13,14,15,16,17-tetradecahydro-1H-cyclopent a[a]phenanthren-3-ol

(3S,8S,9S,10R,13R,14S,17R)-10,13-Diméthyl-17-[(2R)-6-méthyl-2-heptanyl]-2,3,4,7,8,9,10,11,12,13,14,15,16,17-tétradécahydro-1H-cyclopent a[a]phénanthrén-3-ol [French]

(3β)-cholest-5-en-3-ol [ACD/IUPAC Name]

(3β)-Cholest-5-en-3-ol [German] [ACD/IUPAC Name]

(3β)-Cholest-5-én-3-ol [French] [ACD/IUPAC Name]

3b-Hydroxy-5-cholestene

3β-Hydroxycholest-5-ene

5:6-Cholesten-3b-ol

More...

▼ Properties

Experimental data Predicted - ACD/Labs Predicted - EPISuite Predicted - ChemAxon

Data supplied by datasources and users.

• Experimental Physico-Chemical Properties

Experimental Melting Point: ?

149 °C Tokyo Chemical Industry Ltd C0318

147-150 °C Alfa Aesar

148-150 °C Oxford University Chemical Safety Data <http://msds.chem.ox.ac.uk/CH/cholesterol.html>

147-150 °C Alfa Aesar A11470

Experimental Boiling Point: ?

360 °C Alfa Aesar

360 °C Oxford University Chemical Safety Data <http://msds.chem.ox.ac.uk/CH/cholesterol.html>

360 °C Alfa Aesar A11470

Experimental Optical Rotation: ?

-36 Alfa Aesar A11470

Experimental Gravity: ?

1.067 g/mL Alfa Aesar A11470

• Predicted Physico-Chemical Properties

Predicted Melting Point: ?

149 °C Tokyo Chemical Industry Ltd

149 °C Tokyo Chemical Industry Ltd C0318

PubChem

NIH U.S. National Library of Medicine National Center for Biotechnology Information

PubChem OPEN CHEMISTRY DATABASE

Compound Summary for CID 74989

ATOVAQUONE

STRUCTURE VENDORS DRUG INFO PHARMACOLOGY LITERATURE PATENTS BIOACTIVITIES

PubChem CID: 74989

Chemical Names: ATOVAQUONE; Mepron; 95233-18-4; Wellvone; Acugel; 566C80 [More...](#)

Molecular Formula: $C_{22}H_{19}ClO_3$

Molecular Weight: 366.841 g/mol

InChI Key: B5JMWHQBCZFXBR-UHFFFAOYSA-N

Drug Information: [Drug Indication](#) [Therapeutic Uses](#) [Clinical Trials](#) [FDA Orange Book](#) [FDA UNII](#)

Safety Summary: [Laboratory Chemical Safety Summary \(LCSS\)](#)

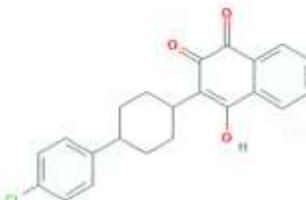
ATOVAQUONE is a hydroxynaphthoquinone that has antimicrobial activity and is being used in antimalarial protocols.

Contents

- 1 2D Structure
- 2 3D Conformer
- 3 Names and Identifiers
- 4 Chemical and Physical Properties
- 5 Related Records
- 6 Chemical Vendors
- 7 Drug and Medication Information
- 8 Pharmacology and Biochemistry
- 9 Use and Manufacturing
- 10 Identification
- 11 Safety and Hazards
- 12 Toxicity
- 13 Literature
- 14 Patents
- 15 Biomolecular Interactions and Pathways
- 16 Biological Test Results
- 17 Classification
- 18 Information Sources

1 2D Structure

Search Download Get Image



Magnify

2 3D Conformer

Search Download Get Image

NCCT Chemistry Dashboard

 United States Environmental Protection Agency

Home Advanced Search Batch Search Lists

Search All Data

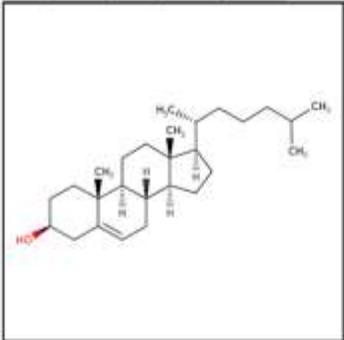
Chemistry Dashboard

Submit Comment Share Copy Aa Aa Aa

Cholesterol

57-88-5 | DTXSID3022401

Searched by Approved Name: Found 1 result for 'cholesterol'.



Wikipedia

Cholesterol, from the Ancient Greek chole- (bile) and stereos (solid) followed by the chemical suffix -ol... [Read more](#)

Intrinsic Properties

Structural Identifiers

Related Compounds

Presence in Lists

Record Information

Chemical Properties Env. Fate/Transport Toxicity Values (Beta) ADME (Beta) Exposure Bioassays

Similar Molecules (Beta) Synonyms Literature External Links Comments

Summary Download as: TSV Excel SDF

Data quality issues

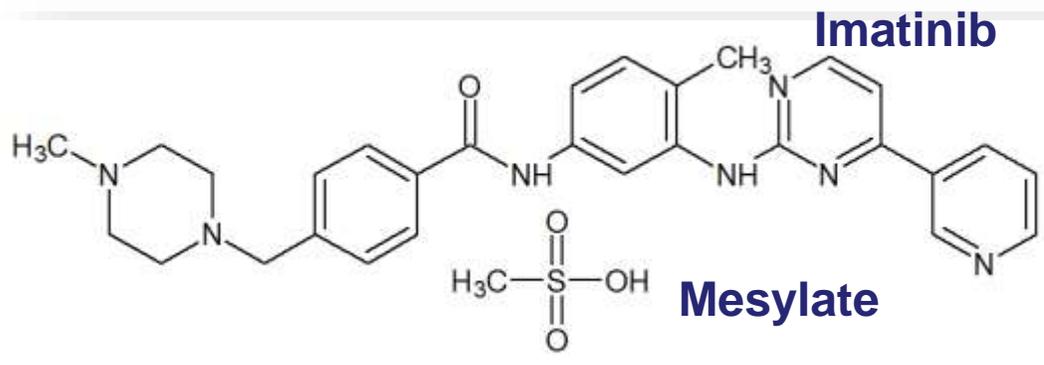
Robochemistry

Proliferation of errors in public and private databases

Automated quality control system

Ambiguities

What Is Gleevec?



The free chemical database

About | More Searches | Web APIs

Gleevec

ChemSpider ID: 11885
Molecular Formula: C₂₉H₃₅N₇O₄S
Average mass: 493.602740
Monoisotopic mass: 493.602740
▼ Systematic name: 4-[[4-(4-methyl-1-piperazinyl)methyl]phenyl]methanesulfonamide

ChemSpider

Structure	<p>Download: MOL SDF SMILES InChI Display: 2D Structure 3D Structure</p>
Synonyms	<ul style="list-style-type: none">Imatinib MesylateImatinib MethanesulfonateSTI-571
Brand names	<ul style="list-style-type: none">GleevecGlivec

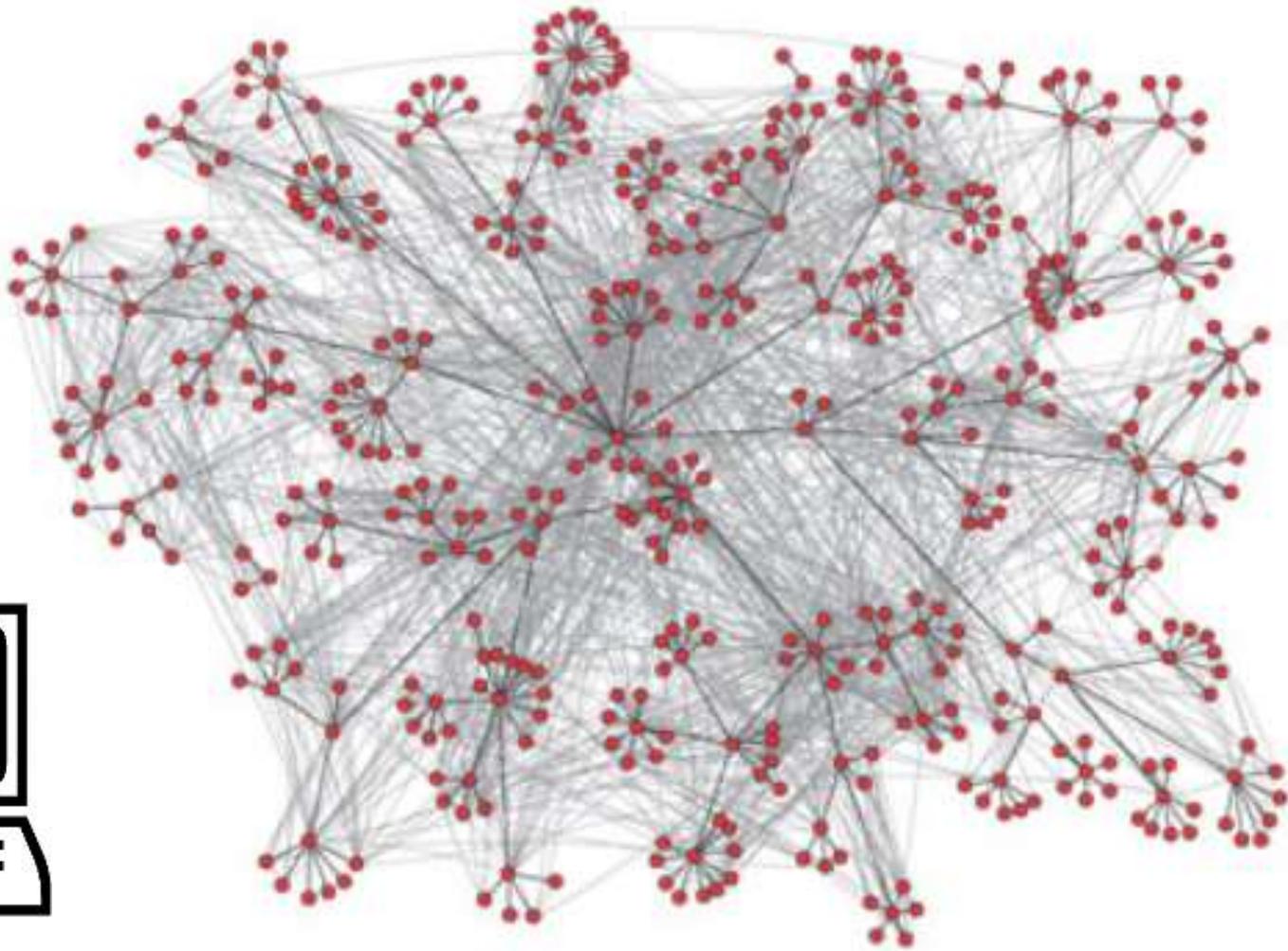
Drugbank

[Imatinib; 152459-95-5; sti-571 ...](#)
MW: 493.602740 g/mol MF: C₂₉H₃₅N₇O₄S
IUPAC name: 4-[[4-(4-methylpiperazin-1-yl)methyl]phenyl]methanesulfonamide
[Active in 205 BioAssays](#) | [Tested in 1376 BioAssays](#)
CID: 5291
[Similar Compounds](#) | [Same Parent, Connectivity](#)
[\(MeSH Keyword\)](#)

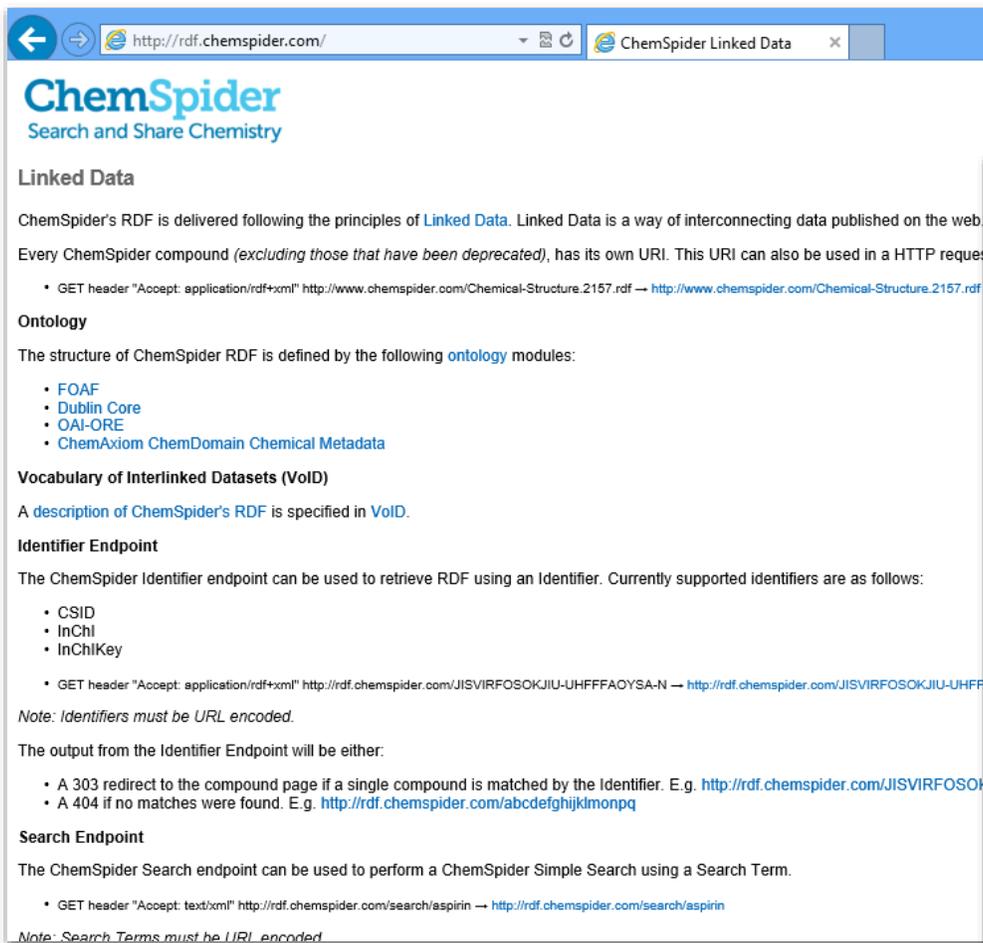
[Imatinib mesylate; Gleevec; Glivec ...](#)
MW: 589.708400 g/mol MF: C₃₀H₃₅N₇O₄S
IUPAC name: methanesulfonic acid; 4-[[4-(4-methylpiperazin-1-yl)methyl]phenyl]methanesulfonamide
[Active in 35 BioAssays](#) | [Tested in 679 BioAssays](#)
CID: 123596
[Similar Compounds](#) | [Same Parent, Connectivity](#)
[\(MeSH Keyword\)](#)

PubChem

We live in a hyperconnected World



What is “sharing in a proper way”?



ChemSpider
Search and Share Chemistry

Linked Data

ChemSpider's RDF is delivered following the principles of [Linked Data](#). Linked Data is a way of interconnecting data published on the web. Every ChemSpider compound (excluding those that have been deprecated), has its own URI. This URI can also be used in a HTTP request:

- GET header "Accept: application/rdf+xml" <http://www.chemspider.com/Chemical-Structure.2157.rdf> → <http://www.chemspider.com/Chemical-Structure.2157.rdf>

Ontology

The structure of ChemSpider RDF is defined by the following [ontology](#) modules:

- [FOAF](#)
- [Dublin Core](#)
- [OAI-ORE](#)
- [ChemAxiom](#) [ChemDomain](#) [Chemical Metadata](#)

Vocabulary of Interlinked Datasets (VoID)

A description of ChemSpider's RDF is specified in [VoID](#).

Identifier Endpoint

The ChemSpider Identifier endpoint can be used to retrieve RDF using an Identifier. Currently supported identifiers are as follows:

- CSID
- InChI
- InChIKey

- GET header "Accept: application/rdf+xml" <http://rdf.chemspider.com/JISVIRFOSOKJIU-UHFFFAOYSA-N> → <http://rdf.chemspider.com/JISVIRFOSOKJIU-UHFFFAOYSA-N>

Note: Identifiers must be URL encoded.

The output from the Identifier Endpoint will be either:

- A 303 redirect to the compound page if a single compound is matched by the Identifier. E.g. <http://rdf.chemspider.com/JISVIRFOSOKJIU-UHFFFAOYSA-N>
- A 404 if no matches were found. E.g. <http://rdf.chemspider.com/abcdefghijklmonpq>

Search Endpoint

The ChemSpider Search endpoint can be used to perform a ChemSpider Simple Search using a Search Term.

- GET header "Accept: text/xml" <http://rdf.chemspider.com/search/aspirin> → <http://rdf.chemspider.com/search/aspirin>

Note: Search Terms must be URL encoded

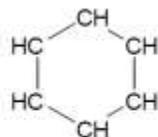
```
<?xml version="1.0" encoding="UTF-8"?>
<!DOCTYPE RDF>
- <rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#" xmlns:chemdomain="http://www.polymerinformatics.com/ChemAxiom/CI"
  xmlns:foaf="http://xmlns.com/foaf/0.1/" xmlns:dcterms="http://purl.org/dc/terms/" xmlns:dc="http://purl.org/dc/elements/1.1/" xmlns:xsd="
  xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#" xml:base="http://www.chemspider.com/Chemical-Structure.123.rdf">
  - <rdf:Description rdf:about="http://www.chemspider.com/">
    <rdfs:label>ChemSpider</rdfs:label>
    <rdfs:seeAlso rdf:resource="http://www.rsc.org/">
  </rdf:Description>
  - <foaf:Document rdf:about="http://www.chemspider.com/Chemical-Structure.123.html">
    <dc:format>text/html</dc:format>
    <dc:rights rdf:resource="http://www.chemspider.com/Disclaimer.aspx">
    <dcterms:creator rdf:resource="http://www.chemspider.com/">
    <dcterms:modified rdf:datatype="http://www.w3.org/2001/XMLSchema#dateTime">2014-03-13</dcterms:modified>
    <ore:aggregatedBy rdf:resource="http://www.chemspider.com/Chemical-Structure.123.rdf#Aggregation"/>
  </foaf:Document>
  - <foaf:Document rdf:about="http://www.chemspider.com/Chemical-Structure.123.rdf">
    <dc:format>application/rdf+xml</dc:format>
    <dc:rights rdf:resource="http://www.chemspider.com/Disclaimer.aspx">
    <dcterms:creator rdf:resource="http://www.chemspider.com/">
    <dcterms:modified rdf:datatype="http://www.w3.org/2001/XMLSchema#dateTime">2014-03-13</dcterms:modified>
    <ore:aggregatedBy rdf:resource="http://www.chemspider.com/Chemical-Structure.123.rdf#Aggregation"/>
    <foaf:primaryTopic rdf:resource="http://www.chemspider.com/Chemical-Structure.123.rdf#Compound"/>
    <foaf:primaryTopic rdf:resource="http://www.chemspider.com/Chemical-Structure.123.rdf#ResourceMap"/>
  </foaf:Document>
  - <ore:Aggregation rdf:about="http://www.chemspider.com/Chemical-Structure.123.rdf#Aggregation">
    <ore:aggregates rdf:resource="http://www.chemspider.com/Chemical-Structure.123.html"/>
    <ore:aggregates rdf:resource="http://www.chemspider.com/Chemical-Structure.123.rdf"/>
    <ore:aggregates rdf:resource="http://www.chemspider.com/Chemical-Structure.123.rdf#Compound"/>
    <ore:aggregates rdf:resource="http://www.chemspider.com/ImagesHandler.ashx?id=123"/>
    <ore:describedBy rdf:resource="http://www.chemspider.com/Chemical-Structure.123.rdf#ResourceMap"/>
  </ore:Aggregation>
  - <chemdomain:NamedChemicalSpecies rdf:about="http://www.chemspider.com/Chemical-Structure.123.rdf#Compound">
    <ore:aggregatedBy rdf:resource="http://www.chemspider.com/Chemical-Structure.123.rdf#Aggregation"/>
    <chemdomain:hasPart rdf:nodeID="autos1"/>
    <foaf:depiction rdf:resource="http://www.chemspider.com/ImagesHandler.ashx?id=123"/>
  </chemdomain:NamedChemicalSpecies>
  - <ore:ResourceMap rdf:about="http://www.chemspider.com/Chemical-Structure.123.rdf#ResourceMap">
    <ore:describes rdf:resource="http://www.chemspider.com/Chemical-Structure.123.rdf#Aggregation"/>
  </ore:ResourceMap>
  - <rdf:Description rdf:about="http://www.chemspider.com/Disclaimer.aspx">
    <dc:language rdf:datatype="http://www.w3.org/2001/XMLSchema#language">en-gb</dc:language>
    <dc:title>Disclaimer</dc:title>
    <dcterms:creator rdf:resource="http://www.chemspider.com/">
    <dcterms:modified rdf:datatype="http://www.w3.org/2001/XMLSchema#dateTime">2014-03-13</dcterms:modified>
  </rdf:Description>
  - <foaf:Image rdf:about="http://www.chemspider.com/ImagesHandler.ashx?id=123">
    <dc:format>image/png</dc:format>
    <dc:rights rdf:resource="http://www.chemspider.com/Disclaimer.aspx">
    <dcterms:creator rdf:resource="http://www.chemspider.com/">
    <dcterms:modified rdf:datatype="http://www.w3.org/2001/XMLSchema#dateTime">2014-03-13</dcterms:modified>
    <ore:aggregatedBy rdf:resource="http://www.chemspider.com/Chemical-Structure.123.rdf#Aggregation"/>
    <foaf:depicts rdf:resource="http://www.chemspider.com/Chemical-Structure.123.rdf#Compound"/>
  </foaf:Image>
```

InChI (<http://www.inchi-trust.org/>)

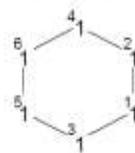
Input Structure



Normalized Structure



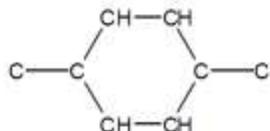
Canonical Numbering



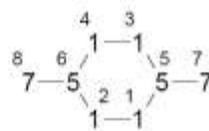
Input Structure



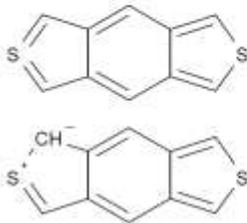
Normalized Structure



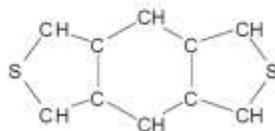
Canonical Numbering



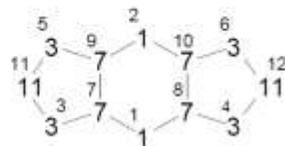
Input Structures



Normalized Structure



Canonical Numbering



```
[InChI version]
1. Main Layer (M):
/(formula)
/c(connections)
/h(H_atoms)
2. Charge Layer
/q(charge)
/p(protons)
3. Stereo Layer
/b(stereo:dbond)
/t(stereo:sp3)
/m(stereo:sp3:inverted)
/s(stereo:type (1=abs, 2=rel, 3=rac))
4. Isotopic Layer (MI):
/i(isotopic:atoms)*
/h(isotopic:exchangeable_H)
/b(isotopic:stereo:dbond)
/t(isotopic:stereo:sp3)
/m(isotopic:stereo:sp3:inverted)
/s(isotopic:stereo:type (1=abs, 2=rel, 3=rac))
5. Fixed H Layer (F):
/f(fixed_H:formula)*
/h(fixed_H:H_fixed)
/q(fixed_H:charge)
/b(fixed_H:stereo:dbond)
/t(fixed_H:stereo:sp3)
/m(fixed_H:stereo:sp3:inverted)
/s(fixed_H:stereo:type (1=abs, 2=rel, 3=rac))
(6.) Fixed/Isotopic Combination (FI)
/i(fixed_H:isotopic:atoms)*
/b(fixed_H:isotopic:stereo:dbond)
/t(fixed_H:isotopic:stereo:sp3)
/m(fixed_H:isotopic:stereo:sp3:inverted)
/s(fixed_H:isotopic:stereo:type (1=abs, 2=rel, 3=rac))
/o(transposition)
```

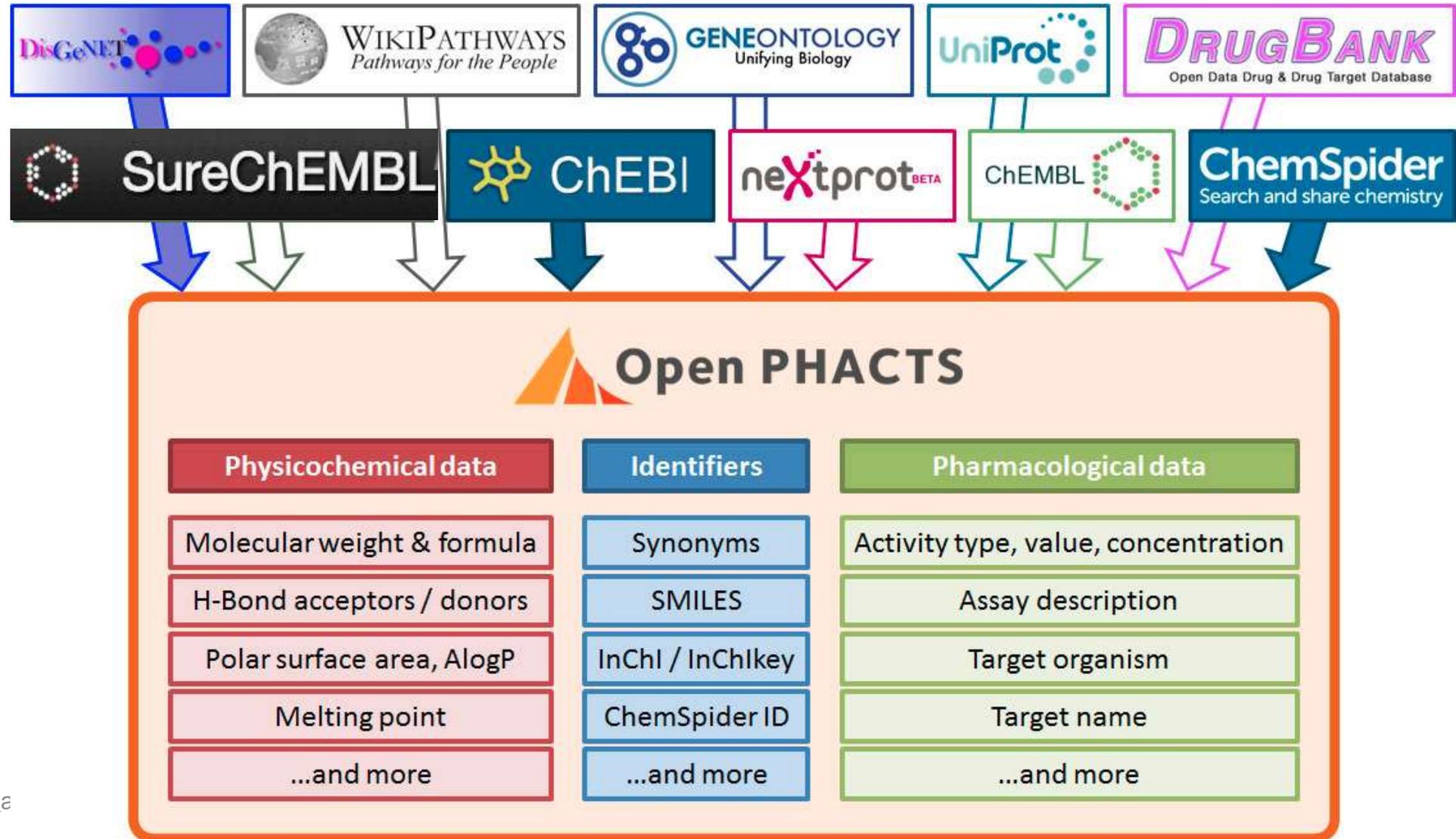
How is this a semantic web problem? Why can't people just be clear?

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.

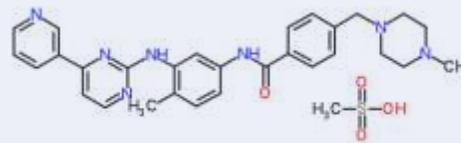
Knowledge is federated



What is lenses?

- 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.

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



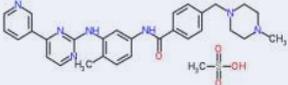
Strict

Relaxed

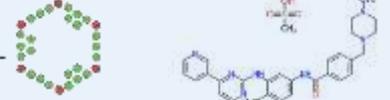
Analysing

Browsing

ChemSpider
Search and share chemistry

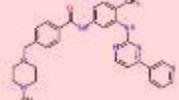


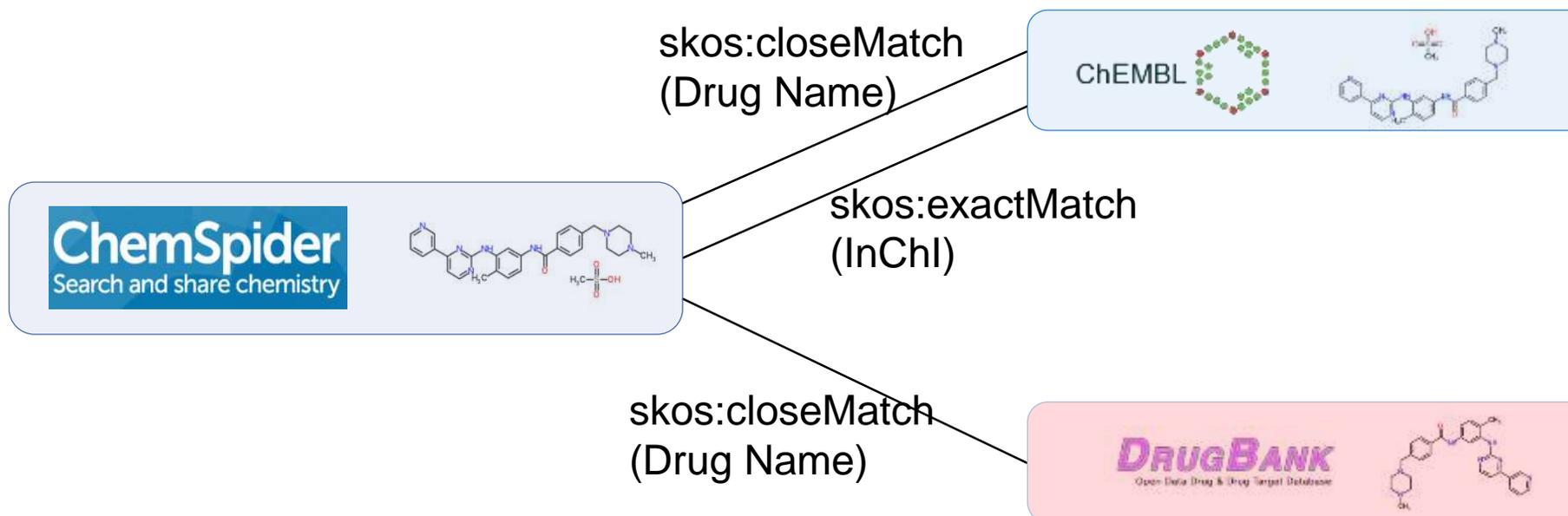
ChEMBL



skos:exactMatch
(InChI)

DRUGBANK
Open-Source Drug & Drug-Target Database





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

Standards



International Union of Pure and Applied Chemistry

Publications

- Technical Reports
- Provisional Recommendations
- IUPAC Books
- E-resources
 - Nomenclature and Terminology
 - Databases
 - Educational resources
- InChI
 - Download
 - R1.01 Summary
 - R1.02beta Summary
 - R1.02 Summary
 - R1.03 Summary
 - Developers, Providers, Publishers
 - News articles
 - Journal articles
- ThermoML
- Chemical Education International

Login | Contact | Forum

→ About → Projects → Publications → Conferences

The IUPAC International Chemical Identifier (InChI)

The IUPAC International Chemical Identifier (InChI™) is a non-proprietary identifier for chemical substances that can be used in printed and electronic data sources thus enabling easier linking of diverse data compilations. It was developed under IUPAC Project 2000-025-1-800 during the period 2000-2004. Details of the project and the history of its progress are available from the [project web site](#).

A list of [software developers](#), [database providers](#), and [journal publishers](#) incorporating InChI in their products is available here.

The InChI™ program version 1.04 is free software developed under the auspices of the International Union of Pure and Applied Chemistry (IUPAC) and the InChI Trust. You can redistribute it and/or modify it under the terms of the [IUPAC-InChI Trust License](#). This is a more permissive version of the [GNU Lesser General Public License version 2.1](#) that was applied to previous versions of the software.

[Download last stable version \(1.04\).](#)

- [InChI 1.01 Software Release Summary](#), August 2006
- [InChI 1.02beta Software Release Summary](#), September 2007
- [InChI Software Version 1.02 – final, implemented for Standard InChI/InChIKey Summary](#), January 2009
- [InChI version 1 \(software version 1.03 for Standard and Non-Standard InChI/InChIKey\) Summary](#), June 2010
- [InChI version 1 \(software version 1.04 for Standard and Non-Standard InChI/InChIKey\)](#), September 2011
[access Software Downloads on www.inchi-trust.org]

IUPAC continues to maintain oversight of InChI development; only systems compliant with the validation protocol (first issued with software release 1.01) are authorised to use the InChI designation. IUPAC welcomes proposals for enhancements. To enable and encourage development of InChI facilities and applications in an Open Source context, a project to encompass this work has been registered with SourceForge.net (see <http://sourceforge.net/projects/inchi>), and people wishing to participate should register via the [SourceForge web page](#).

- News articles on the Identifier are listed [here](#)
- Journal articles that use InChI are listed [here](#)

Online Browsing Platform (OBP)

ISO Search ISO 11238:2012(en)

ISO 11238:2012(en) Health informatics -- Identification of medicinal products -- Data elements and structures for the unique identification and exchange of regulated information on substances

Table of contents

Available in: en

Foreword
Introduction
1. Scope
2. Terms, definitions, symbols and abbreviations
2.1 Terms and definitions
2.2 Symbols and abbreviated terms
3. Requirements
3.1 General
3.2 Concepts required for the unique identification of substances
3.3 Concepts required for the description of substances
3.4 Naming of substances
3.5 Requirements for unique identifier
3.6 Types of substances
3.7 Distinguishing specified substances
Annex A Existing identifiers and molecular structure representations
A.1 Identifiers
A.2 Molecular structure representation
Bibliography

2.1 Terms and definitions

For the purposes of this document, the following terms, definitions and abbreviations apply.

2.1.1 active marker

constituent, or groups of constituents, of an herbal substance, herbal preparation or herbal medicinal product which are of interest for control purposes and are generally accepted to contribute to therapeutic activity

Note 1 to entry: Active markers are not equivalent to analytical or signature markers that serve solely for identification or control purposes.

2.1.2 analytical data

set of elements to describe and capture methods and reference material used to determine purity, potency or identity in a specified substance

2.1.3 chemical bond

condition that occurs when forces acting between two atoms or groups of atoms lead to the formation of a stable discrete molecular entity

2.1.4 chemical substance

type of substance that can be described as a stoichiometric or non-stoichiometric single molecular entity and is not a protein or nucleic acid substance

Note 1 to entry: Chemical substances are generally considered "small" molecules which have associated salts, solvates or ions and may be described using a single definitive or representative structure.

2.1.5 chiral substance

substance whose molecular structure is not superimposable on its mirror image

Figures

2.1.6

[Very incomplete] list of common problems

- Violations of chemical and common sense
- Violations of valence bond theory
- Unsupported format and chemical model features
- Information loss during conversion
- Tautomers
- Stereochemical issues
- Mixtures
- Other classes of chemicals (materials, formulations, biologicals, structurally diverse, etc)
- Equivalence/mapping issues
- Identifiers/names issues
- Etc, etc, etc...

...problems (continued)

- Multiple [historical, proprietary, shortcoming] formats
 - ChemDraw, ChemSketch, AccelrysDraw
 - MOL, SDF
 - SMILES
 - Identifiers
 - Names and Synonyms
- Multiple toolkits/models
 - Open Source (alphabetical)
 - CDK
 - RDKit
 - Indigo
 - OpenBabel
 - Etc...
 - Commercial (alphabetical)
 - CACTVS
 - ChemAxon
 - OpenEye
 - Etc...
- Hystorical software
- No [machine-readable] standards
- ~~No authorities~~ No coordinated efforts!!!

How to link and integrate various resources

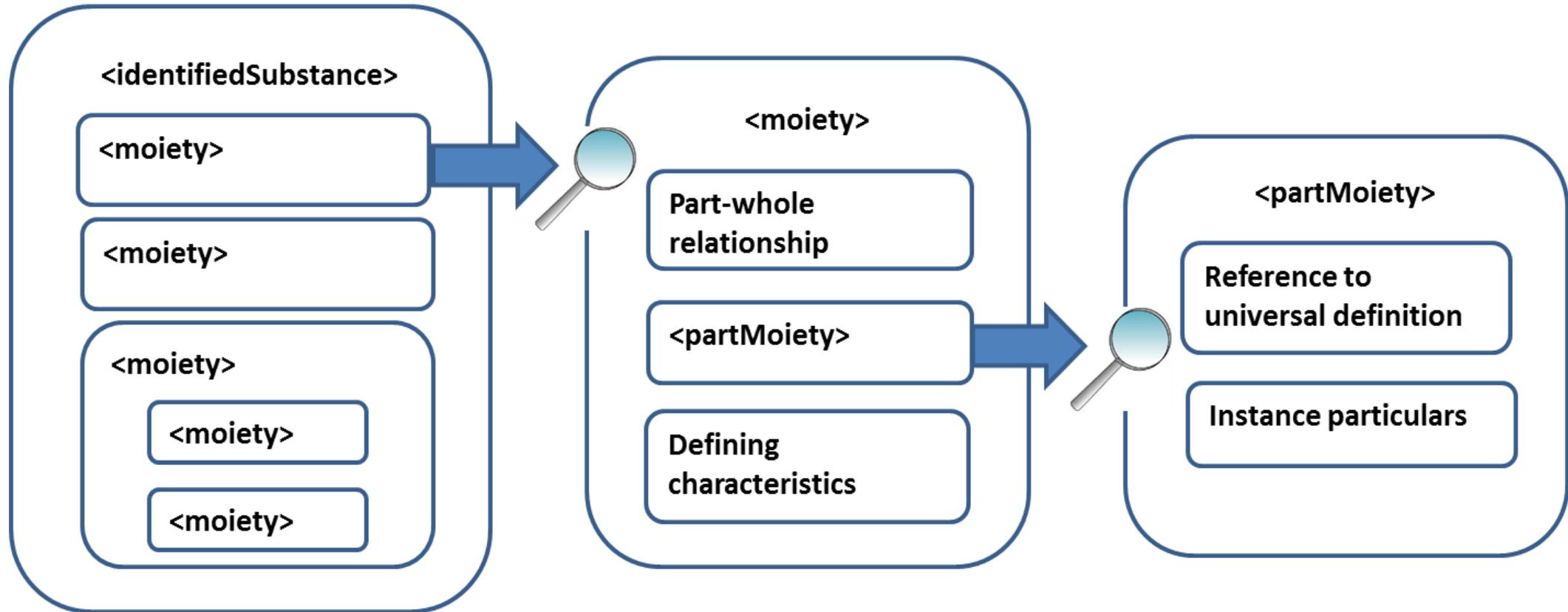
- Available in a variety of databases
- Expressed in a variety of formats
- Some data types are too complex to be exchanged by standard formats. Specific examples are
 - Complex mixtures with defined, or ill-defined concentrations
 - Biological substances
 - Polymers

Structured Product Labeling (SPL)

Health Level Seven (HL7) Structured Product Labeling (SPL)

- an ANSI-accredited data exchange standard
- adopted in 2004 by FDA for the exchange of health and regulatory product and facility data

SPL model



Moiety role	NCIt code	Defining characteristic/representation type	Part-whole relationship	Instance particular	Type	MIME Media Type
Simple chemical	-a	Chemical structure/ MOLFILE, InChI, InChIKey Stereochemistry Type/CV	<quantity>	<id>	Molfile	application/x-mdl-molfile
					InChI	application/x-inchi
					InChIkey	application/x-inchi-key
Protein subunit	C11842 4	Chemical structure/ amino acid letter sequence	<quantity>	<id>	Amino acid sequence	application/x-aa-seq
Polymeric subunit	???	Chemical structure/ MOLFILE, InChI, InChIKey Stereochemistry Type/CV	<quantity>	<id>	DNA Sequence	application/x-dna-seq
					RNA Sequence	application/x-rna-seq

Mixture component	C10324 3	Variable	Letter code	Amino acid
			A (a)	Alanine
Structural modification	C11842 5	Chemical InChIKey Stereochemistry	R (r)	Arginine
			N (n)	Asparagine
			D (d)	Aspartic acid
			B (b)	Asparagine or aspartic acid
			Amino acid connection points	C11842 7
E (e)	Glutamic acid			
Q (q)	Glutamine			
Z (z)	Glutamine or glutamic acid			
G	Glycine			
H (h)	Histidine			
I (i)	Isoleucine			
Linear SRU connection points	???	-	L (l)	Leucine
			K (k)	Lysine
			M (m)	Methionine
			F (f)	Phenylalanine
			P (p)	Proline
			S (s)	Serine
			T (t)	Threonine
			W (w)	Tryptophan
			Y (y)	Tyrosine
			V (v)	Valine
			X	a non-standard amino acid

Stereochemistry type	NCIt code
Square Planar 1 Molecular Geometry	C103211
Square Planar 2 Molecular Geometry	C103212
Square Planar 3 Molecular Geometry	C103213
Square Planar 4 Molecular Geometry	C103214
Tetrahedral Molecular Geometry	C103215
Octahedral 12 Molecular Geometry	C103216
Octahedral 22 Molecular Geometry	C103217
Octahedral 21 Molecular Geometry	C103218
Cahn-Ingold-Prelog Priority System	C103219
Axial R	C103220
Axial S	C103221

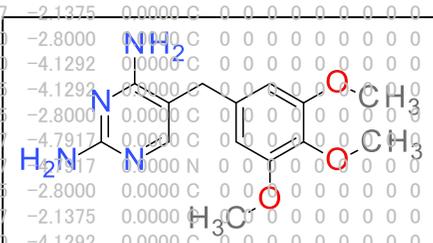
Representation of chemical substance in SPL standard

- Chemical substance

- Chemical structure (MOLFILE)

-FDASRS-04291423352D

```
21 22 0 0 0 0 0 0 0 0999 V2000
2.3000 -2.8000 0.0000 N 0 0 0 0 0 0 0 0 0 0 0 0
3.4417 -2.1375 0.0000 C 0 0 0 0 0 0 0 0 0 0 0 0
4.6000 -2.8000 0.0000 C 0 0 0 0 0 0 0 0 0 0 0 0
2.3000 -4.1292 0.0000 C 0 0 0 0 0 0 0 0 0 0 0 0
9.1875 -4.1292 0.0000 C 0 0 0 0 0 0 0 0 0 0 0 0
9.1875 -2.8000 0.0000 C 0 0 0 0 0 0 0 0 0 0 0 0
8.0417 -4.7917 0.0000 C 0 0 0 0 0 0 0 0 0 0 0 0
3.4417 -4.1292 0.0000 N 0 0 0 0 0 0 0 0 0 0 0 0
6.8875 -2.8000 0.0000 C 0 0 0 0 0 0 0 0 0 0 0 0
5.7417 -2.1375 0.0000 C 0 0 0 0 0 0 0 0 0 0 0 0
4.6000 -4.1292 0.0000 C 0 0 0 0 0 0 0 0 0 0 0 0
8.0417 -2.1375 0.0000 C 0 0 0 0 0 0 0 0 0 0 0 0
6.8875 -4.1292 0.0000 C 0 0 0 0 0 0 0 0 0 0 0 0
3.4417 -0.8125 0.0000 N 0 0 0 0 0 0 0 0 0 0 0 0
```



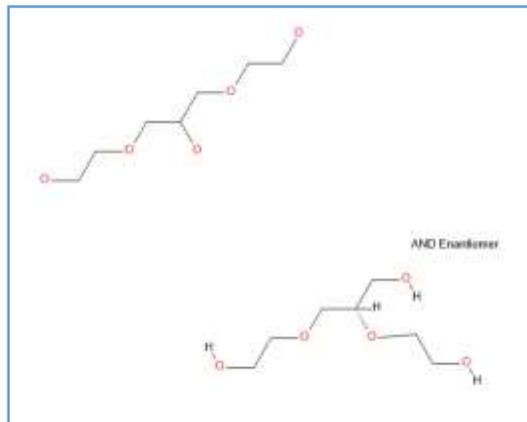
- InChI=1S/C14H18N4O3/c1-19-10-5-8(6-11(20-2)12(10)21-3)4-9-7-17-14(16)18-13(9)15/h5-7H,4H2,1-3H3,(H4,15,16,17,18)
 - IEDVJHCEMCRBQM-UHFFFAOYSA-N

Mixtures

IDMP:

- Mixture substances shall be described as simple combinations of single substances that are either isolated together or are the result of the same synthetic process.
- Mixture substances shall not be combinations of diverse material brought together to form a product.

Mixtures



```
1  
2 Sywx 041716182920 1 1.00000 0.00000 0  
3  
4 I2 11 0 0 0 839 V2000  
5 20.2330 -4.7150 0.0000 0 0 0 0 0 0 0 0 0  
6 19.0499 -4.7150 0.0000 0 0 0 0 0 0 0 0 0  
7 10.4904 -5.7300 0.0000 0 0 0 0 0 0 0 0 0  
8 17.2713 -5.7300 0.0000 0 0 0 0 0 0 0 0 0  
9 16.6877 -6.7615 0.0000 0 0 0 0 0 0 0 0 0  
10 15.5008 -6.7615 0.0000 0 0 0 0 0 0 0 0 0  
11 14.9161 -7.7844 0.0000 0 0 0 0 0 0 0 0 0  
12 15.7350 -7.7844 0.0000 0 0 0 0 0 0 0 0 0  
13 10.0490 -6.7615 0.0000 0 0 0 0 0 0 0 0 0  
14 20.0216 -3.5020 0.0000 0 0 0 0 0 0 0 0 0  
15 22.0027 -3.5020 0.0000 0 0 0 0 0 0 0 0 0  
16 22.5933 -2.5700 0.0000 0 0 0 0 0 0 0 0 0  
17 1 2 1 0 0 0  
18 2 3 1 0 0 0  
19 3 4 1 0 0 0  
20 4 5 1 0 0 0  
21 5 0 1 0 0 0  
22 6 7 1 0 0 0  
23 7 8 1 0 0 0  
24 8 3 1 0 0 0  
25 1 10 1 0 0 0  
26 10 11 1 0 0 0  
27 11 12 1 0 0 0  
28 H END  
29 <MIX_SUBSTANCE_ID>  
30 1794  
31  
32 <MIX_INIT>  
33 00Y855L00  
34  
35 <MIXTURE_TYPE>  
36 ALL  
37  
38 <MIX_DESC_PART1>  
39  
40 <MIX_DESC_PART2>  
41  
42 <MIX_COMMENTS>  
43  
44 <SUBSTANCE_ID>  
45 177122  
46  
47 <UMI1>  
48 90C1P9T912  
49  
50 <DESC_PART1>  
51 H VMD 8000 COLLECTOR  
52 H VMD 8000 SYSTEMS ATRO-C 11  
53 H VMD 800 COLLECTOR  
54 H VMD  
55 <COMMENTS>  
56 <MIX_SUBSTANCE_ID>  
57 1794  
58  
59 <STRUCTURE_ID>  
60 00470  
61  
62 <SUBSTANCE_NAME>  
63 8555
```

```
CHEM  
YHXCXTH047000420 0 0.0000 0.0000 0  
H O O H H 000 0000  
H VMD 8000 C08  
H VMD 8000 C08 20 15 0 0 0 8 80000  
H VMD 8000 C08  
H VMD 1 1 20.2330 -4.7150 0 0  
H VMD 2 1 19.0499 -4.7150 0 0  
H VMD 3 1 10.4904 -5.7300 0 0 0 0 0 0 0  
H VMD 4 1 17.2713 -5.7300 0 0  
H VMD 5 1 16.6877 -6.7615 0 0  
H VMD 6 1 15.5008 -6.7615 0 0  
H VMD 7 1 14.9161 -7.7844 0 0  
H VMD 8 1 15.7350 -7.7844 0 0  
H VMD 9 1 10.0490 -6.7615 0 0  
H VMD 10 1 20.0216 -3.5020 0 0  
H VMD 11 1 22.0027 -3.5020 0 0  
H VMD 12 1 22.5933 -2.5700 0 0  
H VMD 13 1 1 2 1 0 0 0  
H VMD 14 1 2 3 1 0 0 0  
H VMD 15 1 3 4 1 0 0 0  
H VMD 16 1 4 5 1 0 0 0  
H VMD 17 1 5 0 1 0 0 0  
H VMD 18 1 6 7 1 0 0 0  
H VMD 19 1 7 8 1 0 0 0  
H VMD 20 1 8 3 1 0 0 0  
H VMD 21 1 10 1 0 0 0  
H VMD 22 1 10 11 1 0 0 0  
H VMD 23 1 11 12 1 0 0 0  
H VMD 24 1 12 1 0 0 0  
H VMD 25 1 10 000  
H VMD 26 1 10 000  
H VMD 27 1 10 000  
H VMD 28 1 10 000  
H VMD 29 1 10 000  
H VMD 30 1 10 000  
H VMD 31 1 10 000  
H VMD 32 1 10 000  
H VMD 33 1 10 000  
H VMD 34 1 10 000  
H VMD 35 1 10 000  
H VMD 36 1 10 000  
H VMD 37 1 10 000  
H VMD 38 1 10 000  
H VMD 39 1 10 000  
H VMD 40 1 10 000  
H VMD 41 1 10 000  
H VMD 42 1 10 000  
H VMD 43 1 10 000  
H VMD 44 1 10 000  
H VMD 45 1 10 000  
H VMD 46 1 10 000  
H VMD 47 1 10 000  
H VMD 48 1 10 000  
H VMD 49 1 10 000  
H VMD 50 1 10 000  
H VMD 51 1 10 000  
H VMD 52 1 10 000  
H VMD 53 1 10 000  
H VMD 54 1 10 000  
H VMD 55 1 10 000  
H VMD 56 1 10 000  
H VMD 57 1 10 000  
H VMD 58 1 10 000  
H VMD 59 1 10 000  
H VMD 60 1 10 000  
H VMD 61 1 10 000  
H VMD 62 1 10 000  
H VMD 63 1 10 000  
H VMD 64 1 10 000  
H VMD 65 1 10 000  
H VMD 66 1 10 000  
H VMD 67 1 10 000  
H VMD 68 1 10 000  
H VMD 69 1 10 000  
H VMD 70 1 10 000  
H VMD 71 1 10 000  
H VMD 72 1 10 000  
H VMD 73 1 10 000  
H VMD 74 1 10 000  
H VMD 75 1 10 000  
H VMD 76 1 10 000  
H VMD 77 1 10 000  
H VMD 78 1 10 000  
H VMD 79 1 10 000  
H VMD 80 1 10 000  
H VMD 81 1 10 000  
H VMD 82 1 10 000  
H VMD 83 1 10 000  
H VMD 84 1 10 000  
H VMD 85 1 10 000  
H VMD 86 1 10 000  
H VMD 87 1 10 000  
H VMD 88 1 10 000  
H VMD 89 1 10 000  
H VMD 90 1 10 000  
H VMD 91 1 10 000  
H VMD 92 1 10 000  
H VMD 93 1 10 000  
H VMD 94 1 10 000  
H VMD 95 1 10 000  
H VMD 96 1 10 000  
H VMD 97 1 10 000  
H VMD 98 1 10 000  
H VMD 99 1 10 000  
H VMD 100 1 10 000  
H VMD 101 1 10 000  
H VMD 102 1 10 000  
H VMD 103 1 10 000  
H VMD 104 1 10 000  
H VMD 105 1 10 000  
H VMD 106 1 10 000  
H VMD 107 1 10 000  
H VMD 108 1 10 000  
H VMD 109 1 10 000  
H VMD 110 1 10 000  
H VMD 111 1 10 000  
H VMD 112 1 10 000  
H VMD 113 1 10 000  
H VMD 114 1 10 000  
H VMD 115 1 10 000  
H VMD 116 1 10 000  
H VMD 117 1 10 000  
H VMD 118 1 10 000  
H VMD 119 1 10 000  
H VMD 120 1 10 000  
H VMD 121 1 10 000  
H VMD 122 1 10 000  
H VMD 123 1 10 000  
H VMD 124 1 10 000  
H VMD 125 1 10 000  
H VMD 126 1 10 000  
H VMD 127 1 10 000  
H VMD 128 1 10 000  
H VMD 129 1 10 000  
H VMD 130 1 10 000  
H VMD 131 1 10 000  
H VMD 132 1 10 000  
H VMD 133 1 10 000  
H VMD 134 1 10 000  
H VMD 135 1 10 000  
H VMD 136 1 10 000  
H VMD 137 1 10 000  
H VMD 138 1 10 000  
H VMD 139 1 10 000  
H VMD 140 1 10 000  
H VMD 141 1 10 000  
H VMD 142 1 10 000  
H VMD 143 1 10 000  
H VMD 144 1 10 000  
H VMD 145 1 10 000  
H VMD 146 1 10 000  
H VMD 147 1 10 000  
H VMD 148 1 10 000  
H VMD 149 1 10 000  
H VMD 150 1 10 000  
H VMD 151 1 10 000  
H VMD 152 1 10 000  
H VMD 153 1 10 000  
H VMD 154 1 10 000  
H VMD 155 1 10 000  
H VMD 156 1 10 000  
H VMD 157 1 10 000  
H VMD 158 1 10 000  
H VMD 159 1 10 000  
H VMD 160 1 10 000  
H VMD 161 1 10 000  
H VMD 162 1 10 000  
H VMD 163 1 10 000  
H VMD 164 1 10 000  
H VMD 165 1 10 000  
H VMD 166 1 10 000  
H VMD 167 1 10 000  
H VMD 168 1 10 000  
H VMD 169 1 10 000  
H VMD 170 1 10 000  
H VMD 171 1 10 000  
H VMD 172 1 10 000  
H VMD 173 1 10 000  
H VMD 174 1 10 000  
H VMD 175 1 10 000  
H VMD 176 1 10 000  
H VMD 177 1 10 000  
H VMD 178 1 10 000  
H VMD 179 1 10 000  
H VMD 180 1 10 000  
H VMD 181 1 10 000  
H VMD 182 1 10 000  
H VMD 183 1 10 000  
H VMD 184 1 10 000  
H VMD 185 1 10 000  
H VMD 186 1 10 000  
H VMD 187 1 10 000  
H VMD 188 1 10 000  
H VMD 189 1 10 000  
H VMD 190 1 10 000  
H VMD 191 1 10 000  
H VMD 192 1 10 000  
H VMD 193 1 10 000  
H VMD 194 1 10 000  
H VMD 195 1 10 000  
H VMD 196 1 10 000  
H VMD 197 1 10 000  
H VMD 198 1 10 000  
H VMD 199 1 10 000  
H VMD 200 1 10 000  
H VMD 201 1 10 000  
H VMD 202 1 10 000  
H VMD 203 1 10 000  
H VMD 204 1 10 000  
H VMD 205 1 10 000  
H VMD 206 1 10 000  
H VMD 207 1 10 000  
H VMD 208 1 10 000  
H VMD 209 1 10 000  
H VMD 210 1 10 000  
H VMD 211 1 10 000  
H VMD 212 1 10 000  
H VMD 213 1 10 000  
H VMD 214 1 10 000  
H VMD 215 1 10 000  
H VMD 216 1 10 000  
H VMD 217 1 10 000  
H VMD 218 1 10 000  
H VMD 219 1 10 000  
H VMD 220 1 10 000  
H VMD 221 1 10 000  
H VMD 222 1 10 000  
H VMD 223 1 10 000  
H VMD 224 1 10 000  
H VMD 225 1 10 000  
H VMD 226 1 10 000  
H VMD 227 1 10 000  
H VMD 228 1 10 000  
H VMD 229 1 10 000  
H VMD 230 1 10 000  
H VMD 231 1 10 000  
H VMD 232 1 10 000  
H VMD 233 1 10 000  
H VMD 234 1 10 000  
H VMD 235 1 10 000  
H VMD 236 1 10 000  
H VMD 237 1 10 000  
H VMD 238 1 10 000  
H VMD 239 1 10 000  
H VMD 240 1 10 000  
H VMD 241 1 10 000  
H VMD 242 1 10 000  
H VMD 243 1 10 000  
H VMD 244 1 10 000  
H VMD 245 1 10 000  
H VMD 246 1 10 000  
H VMD 247 1 10 000  
H VMD 248 1 10 000  
H VMD 249 1 10 000  
H VMD 250 1 10 000  
H VMD 251 1 10 000  
H VMD 252 1 10 000  
H VMD 253 1 10 000  
H VMD 254 1 10 000  
H VMD 255 1 10 000  
H VMD 256 1 10 000  
H VMD 257 1 10 000  
H VMD 258 1 10 000  
H VMD 259 1 10 000  
H VMD 260 1 10 000  
H VMD 261 1 10 000  
H VMD 262 1 10 000  
H VMD 263 1 10 000  
H VMD 264 1 10 000  
H VMD 265 1 10 000  
H VMD 266 1 10 000  
H VMD 267 1 10 000  
H VMD 268 1 10 000  
H VMD 269 1 10 000  
H VMD 270 1 10 000  
H VMD 271 1 10 000  
H VMD 272 1 10 000  
H VMD 273 1 10 000  
H VMD 274 1 10 000  
H VMD 275 1 10 000  
H VMD 276 1 10 000  
H VMD 277 1 10 000  
H VMD 278 1 10 000  
H VMD 279 1 10 000  
H VMD 280 1 10 000  
H VMD 281 1 10 000  
H VMD 282 1 10 000  
H VMD 283 1 10 000  
H VMD 284 1 10 000  
H VMD 285 1 10 000  
H VMD 286 1 10 000  
H VMD 287 1 10 000  
H VMD 288 1 10 000  
H VMD 289 1 10 000  
H VMD 290 1 10 000  
H VMD 291 1 10 000  
H VMD 292 1 10 000  
H VMD 293 1 10 000  
H VMD 294 1 10 000  
H VMD 295 1 10 000  
H VMD 296 1 10 000  
H VMD 297 1 10 000  
H VMD 298 1 10 000  
H VMD 299 1 10 000  
H VMD 300 1 10 000  
H VMD 301 1 10 000  
H VMD 302 1 10 000  
H VMD 303 1 10 000  
H VMD 304 1 10 000  
H VMD 305 1 10 000  
H VMD 306 1 10 000  
H VMD 307 1 10 000  
H VMD 308 1 10 000  
H VMD 309 1 10 000  
H VMD 310 1 10 000  
H VMD 311 1 10 000  
H VMD 312 1 10 000  
H VMD 313 1 10 000  
H VMD 314 1 10 000  
H VMD 315 1 10 000  
H VMD 316 1 10 000  
H VMD 317 1 10 000  
H VMD 318 1 10 000  
H VMD 319 1 10 000  
H VMD 320 1 10 000  
H VMD 321 1 10 000  
H VMD 322 1 10 000  
H VMD 323 1 10 000  
H VMD 324 1 10 000  
H VMD 325 1 10 000  
H VMD 326 1 10 000  
H VMD 327 1 10 000  
H VMD 328 1 10 000  
H VMD 329 1 10 000  
H VMD 330 1 10 000  
H VMD 331 1 10 000  
H VMD 332 1 10 000  
H VMD 333 1 10 000  
H VMD 334 1 10 000  
H VMD 335 1 10 000  
H VMD 336 1 10 000  
H VMD 337 1 10 000  
H VMD 338 1 10 000  
H VMD 339 1 10 000  
H VMD 340 1 10 000  
H VMD 341 1 10 000  
H VMD 342 1 10 000  
H VMD 343 1 10 000  
H VMD 344 1 10 000  
H VMD 345 1 10 000  
H VMD 346 1 10 000  
H VMD 347 1 10 000  
H VMD 348 1 10 000  
H VMD 349 1 10 000  
H VMD 350 1 10 000  
H VMD 351 1 10 000  
H VMD 352 1 10 000  
H VMD 353 1 10 000  
H VMD 354 1 10 000  
H VMD 355 1 10 000  
H VMD 356 1 10 000  
H VMD 357 1 10 000  
H VMD 358 1 10 000  
H VMD 359 1 10 000  
H VMD 360 1 10 000  
H VMD 361 1 10 000  
H VMD 362 1 10 000  
H VMD 363 1 10 000  
H VMD 364 1 10 000  
H VMD 365 1 10 000  
H VMD 366 1 10 000  
H VMD 367 1 10 000  
H VMD 368 1 10 000  
H VMD 369 1 10 000  
H VMD 370 1 10 000  
H VMD 371 1 10 000  
H VMD 372 1 10 000  
H VMD 373 1 10 000  
H VMD 374 1 10 000  
H VMD 375 1 10 000  
H VMD 376 1 10 000  
H VMD 377 1 10 000  
H VMD 378 1 10 000  
H VMD 379 1 10 000  
H VMD 380 1 10 000  
H VMD 381 1 10 000  
H VMD 382 1 10 000  
H VMD 383 1 10 000  
H VMD 384 1 10 000  
H VMD 385 1 10 000  
H VMD 386 1 10 000  
H VMD 387 1 10 000  
H VMD 388 1 10 000  
H VMD 389 1 10 000  
H VMD 390 1 10 000  
H VMD 391 1 10 000  
H VMD 392 1 10 000  
H VMD 393 1 10 000  
H VMD 394 1 10 000  
H VMD 395 1 10 000  
H VMD 396 1 10 000  
H VMD 397 1 10 000  
H VMD 398 1 10 000  
H VMD 399 1 10 000  
H VMD 400 1 10 000  
H VMD 401 1 10 000  
H VMD 402 1 10 000  
H VMD 403 1 10 000  
H VMD 404 1 10 000  
H VMD 405 1 10 000  
H VMD 406 1 10 000  
H VMD 407 1 10 000  
H VMD 408 1 10 000  
H VMD 409 1 10 000  
H VMD 410 1 10 000  
H VMD 411 1 10 000  
H VMD 412 1 10 000  
H VMD 413 1 10 000  
H VMD 414 1 10 000  
H VMD 415 1 10 000  
H VMD 416 1 10 000  
H VMD 417 1 10 000  
H VMD 418 1 10 000  
H VMD 419 1 10 000  
H VMD 420 1 10 000  
H VMD 421 1 10 000  
H VMD 422 1 10 000  
H VMD 423 1 10 000  
H VMD 424 1 10 000  
H VMD 425 1 10 000  
H VMD 426 1 10 000  
H VMD 427 1 10 000  
H VMD 428 1 10 000  
H VMD 429 1 10 000  
H VMD 430 1 10 000  
H VMD 431 1 10 000  
H VMD 432 1 10 000  
H VMD 433 1 10 000  
H VMD 434 1 10 000  
H VMD 435 1 10 000  
H VMD 436 1 10 000  
H VMD 437 1 10 000  
H VMD 438 1 10 000  
H VMD 439 1 10 000  
H VMD 440 1 10 000  
H VMD 441 1 10 000  
H VMD 442 1 10 000  
H VMD 443 1 10 000  
H VMD 444 1 10 000  
H VMD 445 1 10 000  
H VMD 446 1 10 000  
H VMD 447 1 10 000  
H VMD 448 1 10 000  
H VMD 449 1 10 000  
H VMD 450 1 10 000  
H VMD 451 1 10 000  
H VMD 452 1 10 000  
H VMD 453 1 10 000  
H VMD 454 1 10 000  
H VMD 455 1 10 000  
H VMD 456 1 10 000  
H VMD 457 1 10 000  
H VMD 458 1 10 000  
H VMD 459 1 10 000  
H VMD 460 1 10 000  
H VMD 461 1 10 000  
H VMD 462 1 10 000  
H VMD 463 1 10 000  
H VMD 464 1 10 000  
H VMD 465 1 10 000  
H VMD 466 1 10 000  
H VMD 467 1 10 000  
H VMD 468 1 10 000  
H VMD 469 1 10 000  
H VMD 470 1 10 000  
H VMD 471 1 10 000  
H VMD 472 1 10 000  
H VMD 473 1 10 000  
H VMD 474 1 10 000  
H VMD 475 1 10 000  
H VMD 476 1 10 000  
H VMD 477 1 10 000  
H VMD 478 1 10 000  
H VMD 479 1 10 000  
H VMD 480 1 10 000  
H VMD 481 1 10 000  
H VMD 482 1 10 000  
H VMD 483 1 10 000  
H VMD 484 1 10 000  
H VMD 485 1 10 000  
H VMD 486 1 10 000  
H VMD 487 1 10 000  
H VMD 488 1 10 000  
H VMD 489 1 10 000  
H VMD 490 1 10 000  
H VMD 491 1 10 000  
H VMD 492 1 10 000  
H VMD 493 1 10 000  
H VMD 494 1 10 000  
H VMD 495 1 10 000  
H VMD 496 1 10 000  
H VMD 497 1 10 000  
H VMD 498 1 10 000  
H VMD 499 1 10 000  
H VMD 500 1 10 000  
H VMD 501 1 10 000  
H VMD 502 1 10 000  
H VMD 503 1 10 000  
H VMD 504 1 10 000  
H VMD 505 1 10 000  
H VMD 506 1 10 000  
H VMD 507 1 10 000  
H VMD 508 1 10 000  
H VMD 509 1 10 000  
H VMD 510 1 10 000  
H VMD 511 1 10 000  
H VMD 512 1 10 000  
H VMD 513 1 10 000  
H VMD 514 1 10 000  
H VMD 515 1 10 000  
H VMD 516 1 10 000  
H VMD 517 1 10 000  
H VMD 518 1 10 000  
H VMD 519 1 10 000  
H VMD 520 1 10 000  
H VMD 521 1 10 000  
H VMD 522 1 10 000  
H VMD 523 1 10 000  
H VMD 524 1 10 000  
H VMD 525 1 10 000  
H VMD 526 1 10 000  
H VMD 527 1 10 000  
H VMD 528 1 10 000  
H VMD 529 1 10 000  
H VMD 530 1 10 000  
H VMD 531 1 10 000  
H VMD 532 1 10 000  
H VMD 533 1 10 000  
H VMD 534 1 10 000  
H VMD 535 1 10 000  
H VMD 536 1 10 000  
H VMD 537 1 10 000  
H VMD 538 1 10 000  
H VMD 539 1 10 000  
H VMD 540 1 10 000  
H VMD 541 1 10 000  
H VMD 542 1 10 000  
H VMD 543 1 10 000  
H VMD 544 1 10 000  
H VMD 545 1 10 000  
H VMD 546 1 10 000  
H VMD 547 1 10 000  
H VMD 548 1 10 000  
H VMD 549 1 10 000  
H VMD 550 1 10 000  
H VMD 551 1 10 000  
H VMD 552 1 10 000  
H VMD 553 1 10 000  
H VMD 554 1 10 000  
H VMD 555 1 10 000  
H VMD 556 1 10 000  
H VMD 557 1 10 000  
H VMD 558 1 10 000  
H VMD 559 1 10 000  
H VMD 560 1 10 000  
H VMD 561 1 10 000  
H VMD 562 1 10 000  
H VMD 563 1 10 000  
H VMD 564 1 10 000  
H VMD 565 1 10 000  
H VMD 566 1 10 000  
H VMD 567 1 10 000  
H VMD 568 1 10 000  
H VMD 569 1 10 000  
H VMD 570 1 10 000  
H VMD 571 1 10 000  
H VMD 572 1 10 000  
H VMD 573 1 10 000  
H VMD 574 1 10 000  
H VMD 575 1 10 000  
H VMD 576 1 10 000  
H VMD 577 1 10 000  
H VMD 578 1 10 000  
H VMD 579 1 10 000  
H VMD 580 1 10 000  
H VMD 581 1 10 000  
H VMD 582 1 10 000  
H VMD 583 1 10 000  
H VMD 584 1 10 000  
H VMD 585 1 10 000  
H VMD 586 1 10 000  
H VMD 587 1 10 000  
H VMD 588 1 10 000  
H VMD 589 1 10 000  
H VMD 590 1 10 000  
H VMD 591 1 10 000  
H VMD 592 1 10 000  
H VMD 593 1 10 000  
H VMD 594 1 10 000  
H VMD 595 1 10 000  
H VMD 596 1 10 000  
H VMD 597 1 10 000  
H VMD 598 1 10 000  
H VMD 599 1 10 000  
H VMD 600 1 10 000  
H VMD 601 1 10 000  
H VMD 602 1 10 000  
H VMD 603 1 
```

Example Complex Mixtures

- “Aroclors” are complex mixtures of polychlorinated biphenyls (PCBs). There are 209 possible PCBs and different Aroclors are combinations of a series of these 209 variants and at specific ranges of concentrations.
- Ideally SPL will carry information about the individual components and the concentration of each component for a specific Aroclor
- Work in progress and looking promising!

Substances in products

- Small molecules
- Proteins
- Nucleic acids
- Polymers
- Organisms
- Parts of organisms
- Mixtures

FAIR Data Principles

To be Findable:

- F1. (meta)data are assigned a globally unique and eternally persistent identifier.
- F2. data are described with rich metadata.
- F3. (meta)data are registered or indexed in a searchable resource.
- F4. metadata specify the data identifier.

To be Accessible:

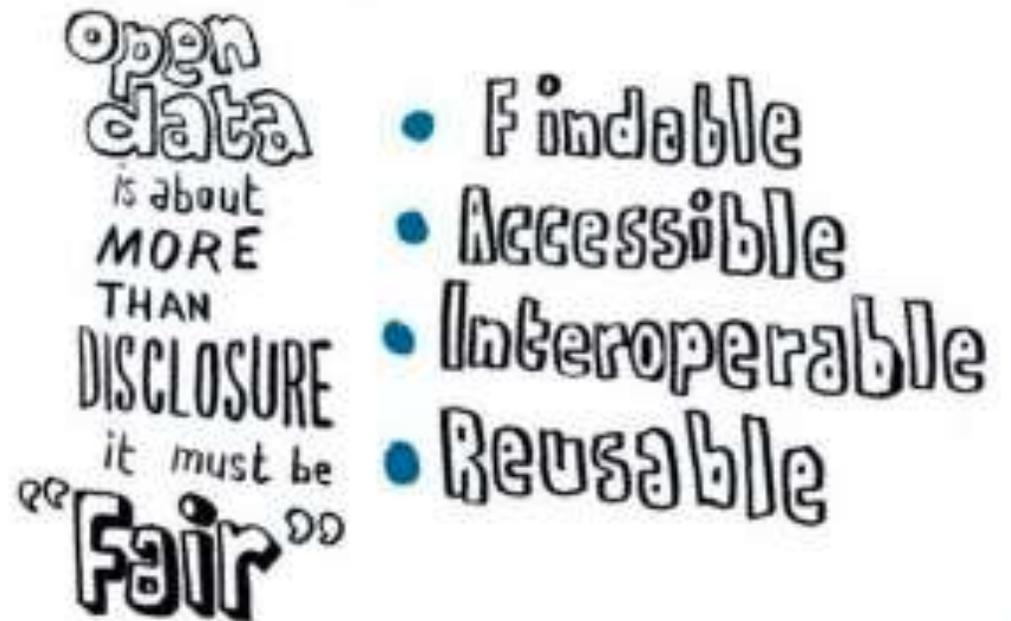
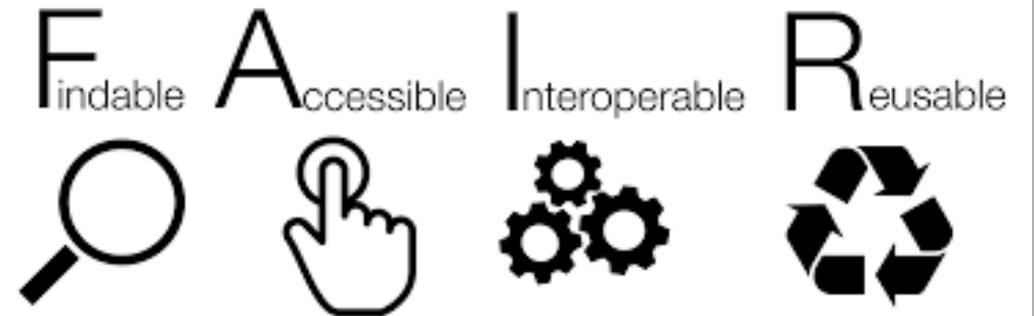
- A1 (meta)data are retrievable by their identifier using a standardized communications protocol.
- A1.1 the protocol is open, free, and universally implementable.
- A1.2 the protocol allows for an authentication and authorization procedure, where necessary.
- A2 metadata are accessible, even when the data are no longer available.

To be Interoperable:

- I1. (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.
- I2. (meta)data use vocabularies that follow FAIR principles.
- I3. (meta)data include qualified references to other (meta)data.

To be Re-usable:

- R1. meta(data) have a plurality of accurate and relevant attributes.
- R1.1. (meta)data are released with a clear and accessible data usage license.
- R1.2. (meta)data are associated with their provenance.
- R1.3. (meta)data meet domain-relevant community standards.



Open Science Data Repository powered by Dataledger™

OPEN SCIENCE DATA REPOSITORY
Powered by Dataledger™

Home Organize wait

ONE PLACE TO STORE YOUR DATA

30M
Chemical Structures

103
Data Sources

UPLOAD AND ORGANIZE
You can upload your files, transfer them from Google Drive, Box.com, DropBox.com etc. And - manage your data your own way!

MACHINE LEARNING
Data models, algorithms and pipelines for cheminformatics and drug discovery.

SHARE & ANNOTATE
We support widely used vocabularies. Plus, you can add your own. Share your work with others.

Powered by Dataledger Technology. ©2017 - Science Data Software, LLC
Development version 0.0.1 build 1, 12.02.2017, 16:14:24



Chemical processing

- Support for chemical formats
- Chemistry validation and standardization
- Automatic processing and visualization

The screenshot displays the Open Science Data Repository interface, powered by Dataledger. The header includes the logo, the text "OPEN SCIENCE DATA REPOSITORY", and navigation links for "Home", "Organize", and a user profile "valt".

The left sidebar lists various data types with their respective counts:

- Articles: 1233
- Images: 10000
- Research: 200
- Structures: 30012645
- Crystals: 86659
- Reactions: 3144655
- Spectra: 6094
- Datasets: 2133

The main content area is titled "DRAFTS / QSAR" and features a grid of 15 chemical structures, each with a corresponding data file name:

- data_Hemeoxyg...
- data_Humancox...
- data_lipinski.mo...
- data_Adrenergic...
- data_Alcoholde...
- data_Arachidon...
- data_Cytochro...
- data_Hemeoxyg...
- data_Humancox...
- data_LaCrossevi...
- data_Nitric_oxid...
- data_Pseudolysi...
- data_Serotonin...
- data_Tryptopha...
- data_lipinski.sdf

Possible solution

- Agreed and machine-readable (digital) standards
- Open-source (and therefore fully transparent) solution
- Organizational AND community support and involvement
- Accessible solution
- Data triaging at data repositories level
- Real-time validation/standardization (API, library, “docker”, etc)

Thank you!

On Web:

scidatasoft.com

Contact us:

info@scidatasoft.com

Slides:

<https://www.slideshare.net/valerytkachenko16>