ontochem

Chemical Ontologies

ChemAxon UGM May 23, 2012

OntoChem GmbH
Heinrich-Damerow-Str. 4
06120 Halle (Saale)
Germany
Tel. +49 345 4780472
Fax: +49 345 4780471
mail: info(at)ontochem.com

Chemical Ontologies - Why ?

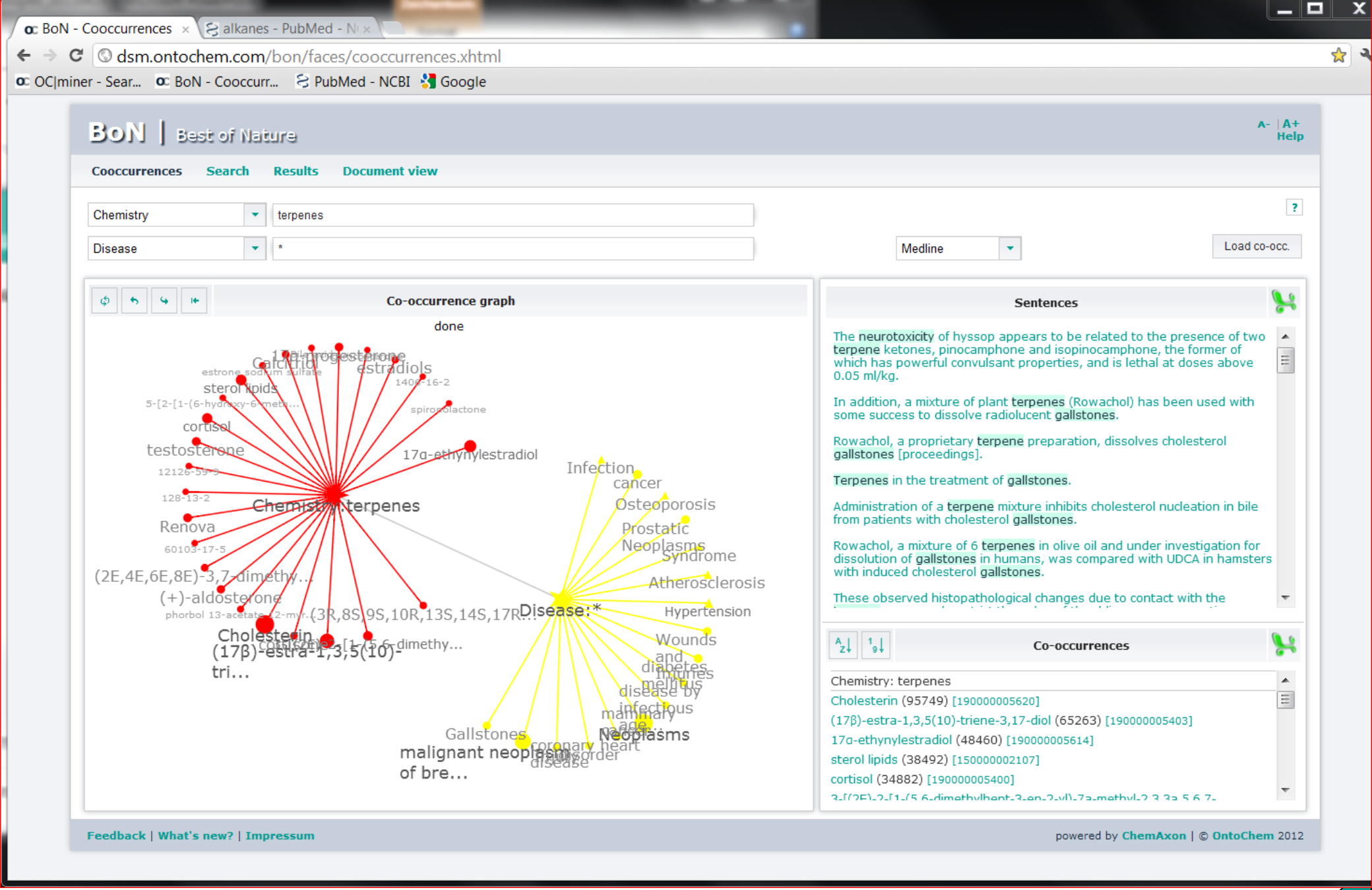
Chemistry has a long history in specialized databases

Compound databases
Reaction databases
Structure-Activity Relationships (SAR)

Open needs

- Smart chemistry search for non-chemists
- Structure search over text documents
- Searches over patents
- Complex high-level chemistry related queries

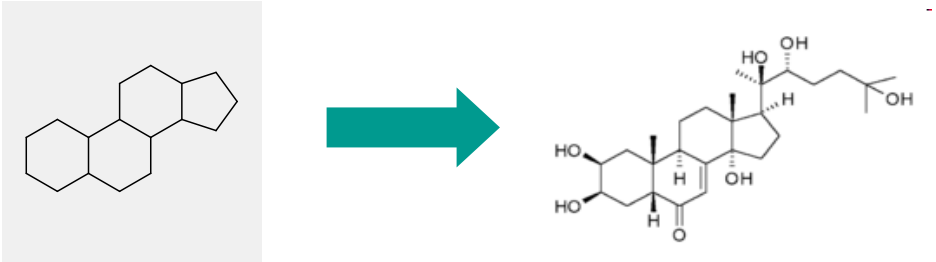
→ **Chemical Ontologies**


BoN - Cooccurrences
alkanes - PubMed
dsm.ontochem.com/bon/faces/cooccurrences.xhtml
OC|miner - Sear...
BoN - Cooccurr...
PubMed - NCBI
Google
BoN | Best of Nature
A- | A+
Help
Cooccurrences
Search
Results
Document view
Chemistry
terpenes
Disease
*
Medline
Load co-occ.
Co-occurrence graph
done
Calcitriol
estrone sodium sulfate
estradiols
sterol lipids
spironolactone
cortisol
testosterone
17α-ethynylestradiol
Chemistry:terpenes
Renova
(2E,4E,6E,8E)-3,7-dimethy...
(+)-aldosterone
Cholesterin
(17β)-estra-1,3,5(10)-tri...
Disease:*
Infection
cancer
Osteoporosis
Prostatic Neoplasms
Syndrome
Atherosclerosis
Hypertension
Wounds and
diabetes
disease by infectious
mammary
Gallstones
Neoplasms
malignant neoplasm of bre...
Sentences
The neurotoxicity of hyssop appears to be related to the presence of two terpene ketones, pinocamphone and isopinocamphone, the former of which has powerful convulsant properties, and is lethal at doses above 0.05 ml/kg.
In addition, a mixture of plant terpenes (Rowachol) has been used with some success to dissolve radiolucent gallstones.
Rowachol, a proprietary terpene preparation, dissolves cholesterol gallstones [proceedings].
Terpenes in the treatment of gallstones.
Administration of a terpene mixture inhibits cholesterol nucleation in bile from patients with cholesterol gallstones.
Rowachol, a mixture of 6 terpenes in olive oil and under investigation for dissolution of gallstones in humans, was compared with UDCA in hamsters with induced cholesterol gallstones.
These observed histopathological changes due to contact with the
Co-occurrences
Chemistry: terpenes
Cholesterin (95749) [190000005620]
(17β)-estra-1,3,5(10)-triene-3,17-diol (65263) [190000005403]
17α-ethynylestradiol (48460) [190000005614]
sterol lipids (38492) [150000002107]
cortisol (34882) [190000005400]
Feedback | What's new? | Impressum
powered by ChemAxon | © OntoChem 2012

Chemical Ontologies - Why ?

Bringing chemistry to non-chemists !

- Search for “steroids” and find “ecdysterone”



- Resolve Markush structures: 1-cycloalkyl isoquinolines

- High-level queries:
  - what diseases can be treated with terpenes ?
  - which proteins are inhibited by terpenes ?
  - which steroids are found in plants ?
  - which compound classes have antibacterial properties ?
  - drug repurposing: alternative uses of my compound ?

Chemical Ontologies – Status Today

PubMed

- PubMed assigns compounds to classes
- 9,097 classes and compounds
- 58,822 synonym terms
- Manual assignment

ChEBI (Chemical Entities of Biological Interest)

- ChEBI assigns compounds to classes
- 30,935 classes and compounds
- 182,608 synonym terms
- Manual assignment

Problems of manual creation and assignment

- Redundant relationships
- Homonyms in the hierarchy path
- Missing relationships
- Wrong relationships

Chemical Ontologies – Status Today

Ontology Problems (e.g. MeSH, ChEBI)

- **Redundant relationships**
  Amino Acids <- Threonine
  Amino Acids <- Essential Amino Acids <- Threonine
  Amino Acids <- Neutral Amino Acids <- Threonine

- **Homonyms in the hierarchy path**
  Insulin (regular insulin) <- regular pork insulin (regular insulin)

- **Missing relationships**
  Histidine -> imidazole, carboxylic acid, amine
  Steroids -> terpenes

- **Wrong relationships**
  Chlorohydrocarbon <- 2,2-bis(4-chlorophenyl)ethanol

Chemical Ontologies – Status Today

Conceptual Problems :

Chemistry class terms in life science

- derivative, backbone, scaffold, compound class, compound
- substructure, superstructure, tautomer, diastereomer, enantiomer, epimer …

ChEBI

- *is_conjugate_base_of, is_conjugate_acid_of*
- *is_tautomer_of, has_parent_hydride, is_enantiomer_of*
- *has_functional_parent, is_substituent_group_from, has_role*

… common understanding of terms needed !

Chemical Ontology Editor

Goal: Create an ontology editor for chemistry that allows

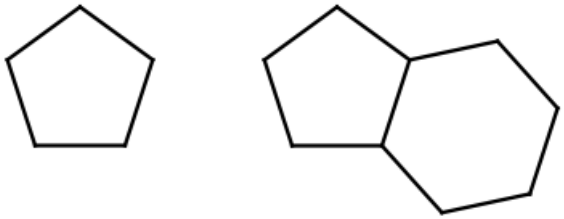
- Classify chemical compounds into structure classes, e.g. steroids, terpenes, alcohols etc.
- Automated assignment of compound classes to compounds in structure files (SDF, SMI or others) or in databases
- Quality checks:
  - loop detection,
  - redundant links,
  - homonyms,
  - chemical validity and accessibility
- Optimized for named entity recognition in text:
  - recognize compounds and compound classes, e.g. propane, propanes, propane derivatives and propyl substituent

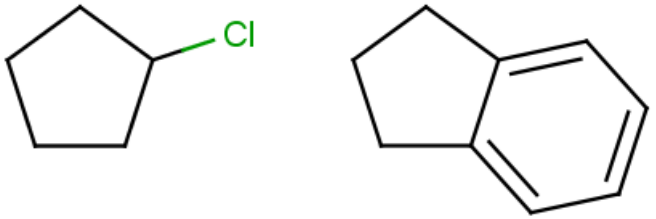
Chemical Ontologies – Design Principles

"*is_a*" compound class

is_a cycloalkane, cycloalkanes: **not** *is_a* cycloalkane

defined by common (biological) understanding

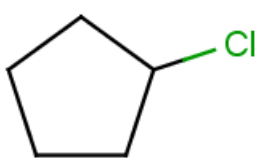


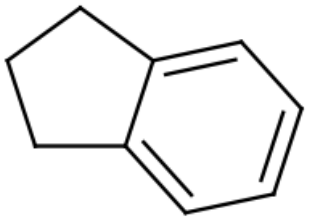


"*is_a*" compound derivative class

cycloalkane derivatives (contains cycloalkane):

cycloalkane derivatives, scaffold, substituent, group, moiety

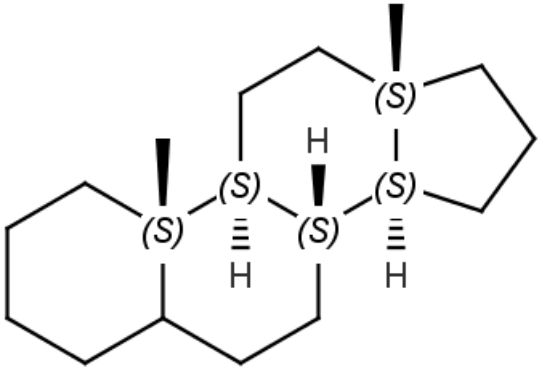


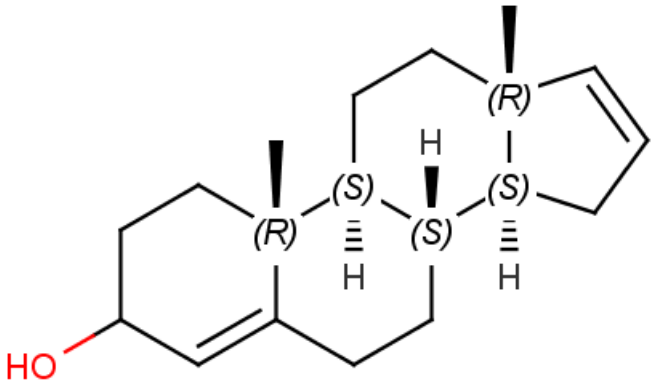


Chemical Ontologies – Design Principles

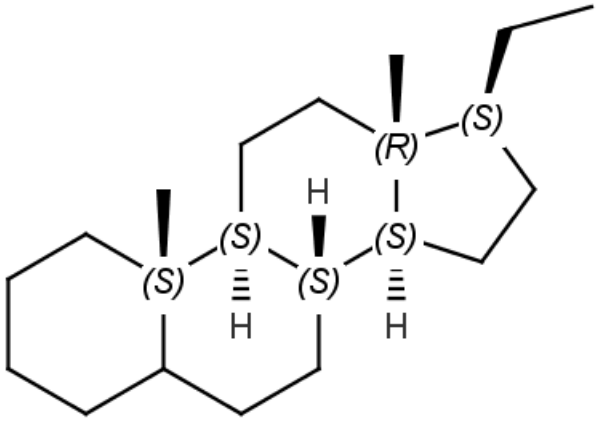
"*is_a*" class type

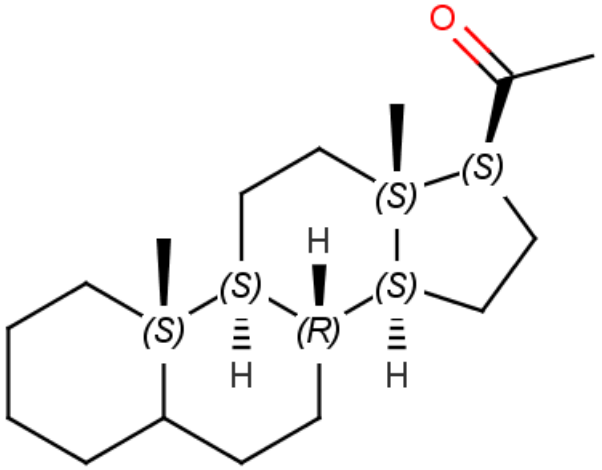
is_a androstane, **not** *is_a* pregnane





is_a pregnane, **not** *is_a* androstane





Chemical Ontologies – Design Principles

"*is_a*" class type

organic compounds

lipids

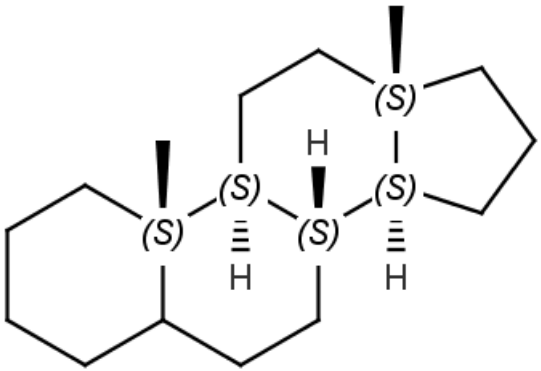
prenol lipids

terpenes

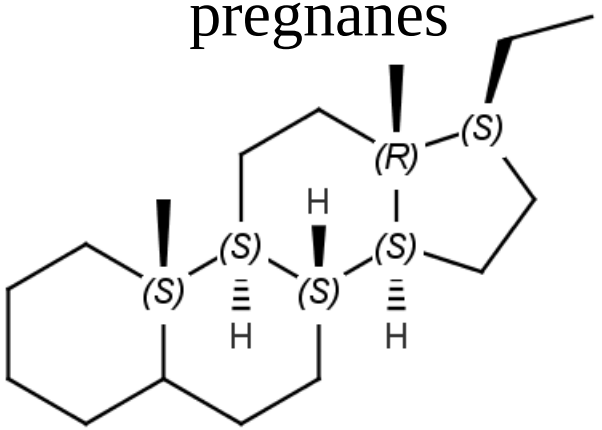
triterpenes

steroids

androstanes



pregnanes



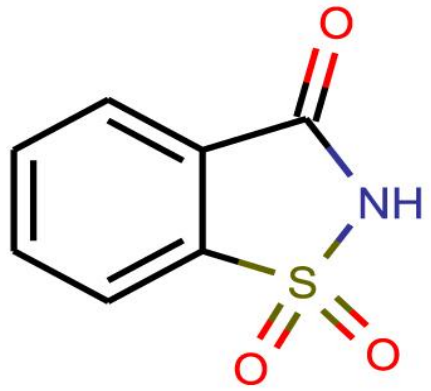
SMILES

Compounds have

SMILES = Simplified Molecular Input Line Entry Specification

- structures as ASCII strings
- human readable
- 2 symbol types: Atoms (C,O,N,S) and bonds (-,=,#,:)

SMILES:
O=C1NS(=O)(=O)C2=CC=CC=C12



→ SMILES for describing compounds

SMARTS

Compound classes may have

SMARTS = SMiles Arbitrary Target Specification

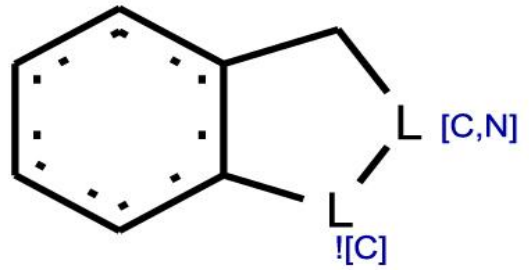
- molecular pattern
- atom lists and non-lists ([C,N], ![C])
- bond lists: any, single OR double, double OR aromatic

Compound classes may not have SMARTS,

but only children with SMARTS

e.g. “lipids”, “terpenes”

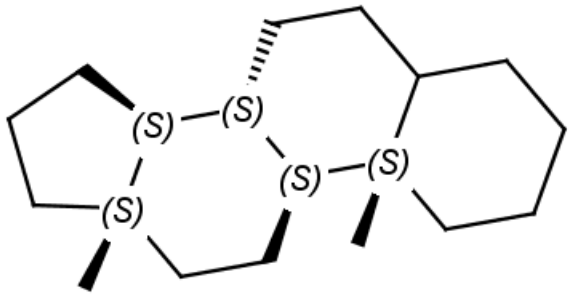
SMARTS:
[#6]1[#6][#6][#6]2[!#6][C,N]C[#6]2[#6]1



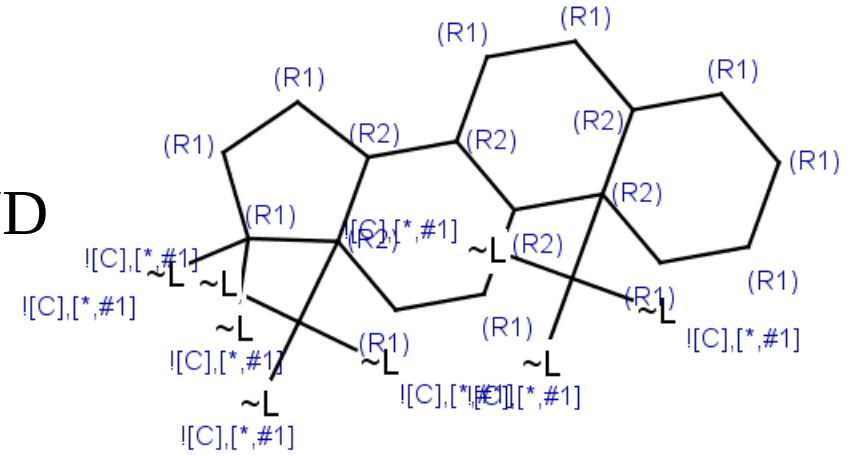
→ SMARTS for describing compound classes

Chemical Ontologies – Design Principles

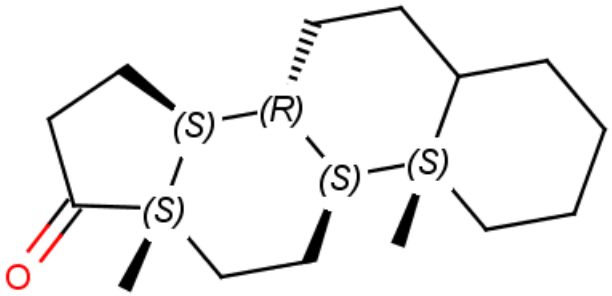
class definition is using SMARTS logic: androstanes



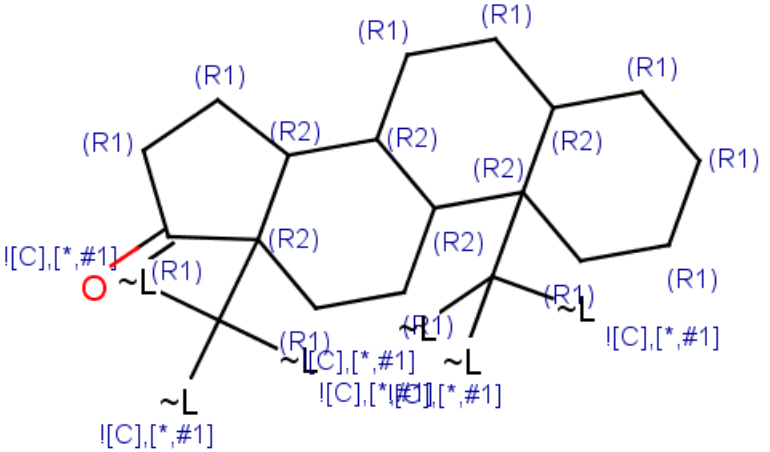
AND



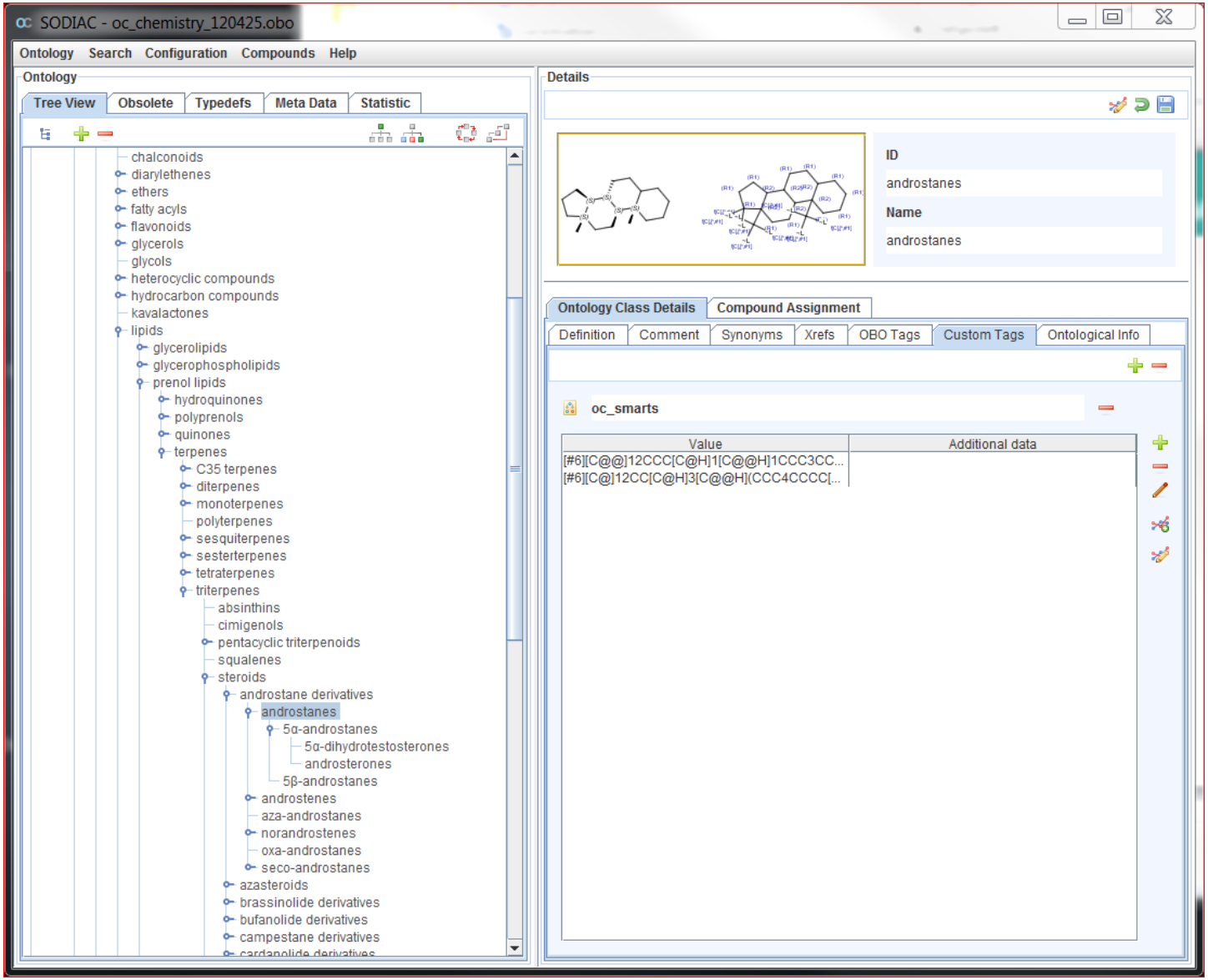
OR



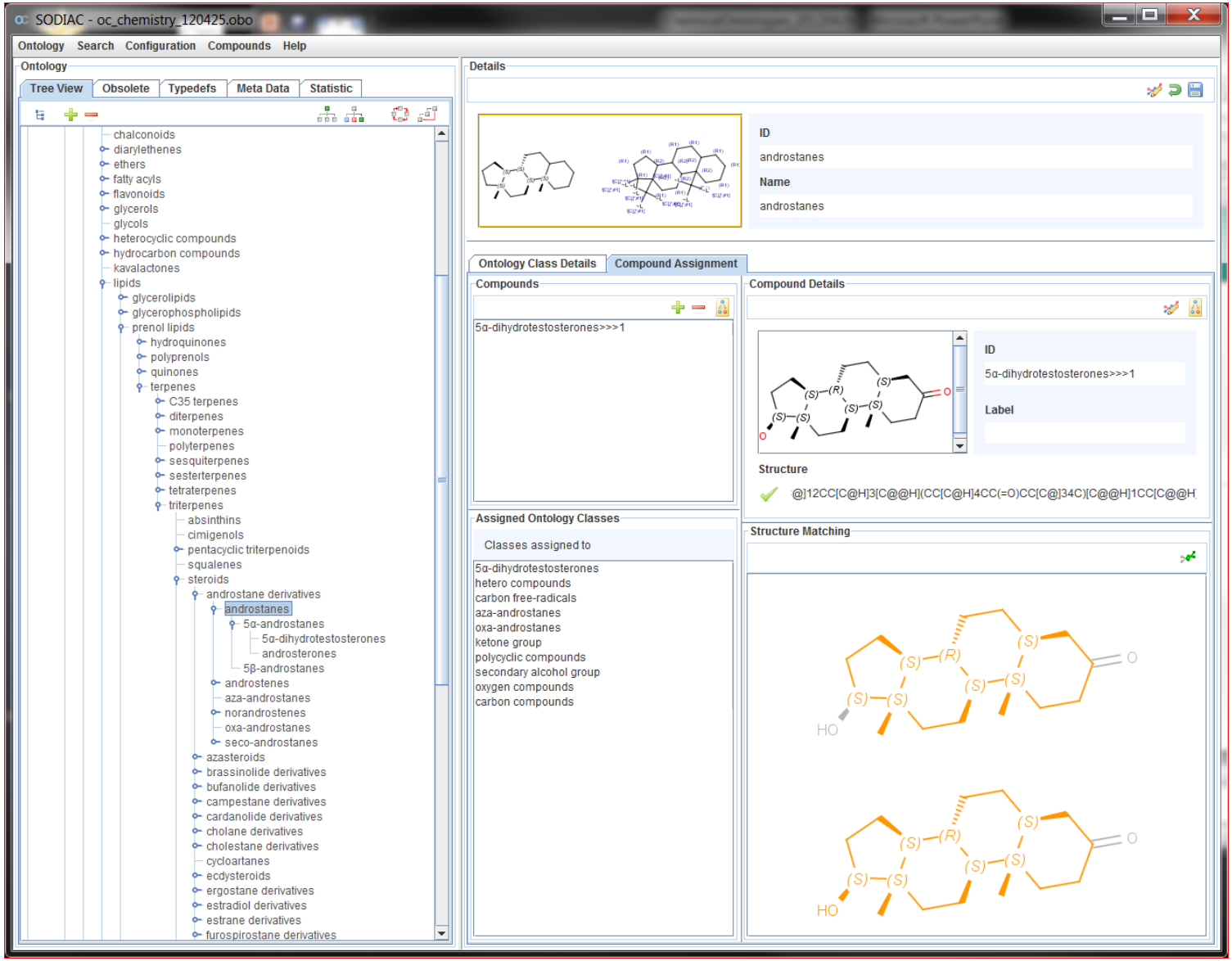
AND



Chemical Ontologies Editor - SODIAC



Chemical Ontologies Editor - SODIAC



Chemical Ontologies Editor - SODIAC

- reads SMILES, SDF, OWL and OBO files
- writes SMILES, SDF and OBO files

- chemistry ontologies based on SMARTS
- integrates with name2structure tools
- allows to edit and export any OBO ontology
- integrity and ontology checking
- structure checking

- assigns compound classes in files or databases

Examples

PubChem:

- **37 million compounds**
- **currently found 337 million compound class annotations with SODIAC**

Chemistry in Medline:

- found 92.813 chemicals in about 19 million abstracts
  - 1.225.000 times water...
  - 1.191.000 times calcium
  - 756.800 times glucose
  - 353.000 times cholesterol
- natural products - the largest compound class with interesting data

Firefox

OC|miner - Document View

oc36:8080/ocm/showDocument.jsp?docid=8&query=water&rep=http%3A%2F%2F192.168.0.36%3A9941%2F_dia

Google

Meistbesuchte Seiten | BoN - Cooccurrences | Efficacy | Lesezeichen

compound group

ontochem

OC|miner

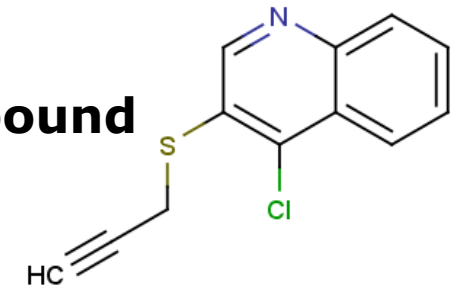
water | Search | diamond (http://192.168.0.36:9941/) | Hits/page: 10

Show parsed query | Simple Syntax

anaphora compound

compound class

OPSIN compound



lipopni ty.

As an extension of our work on the development of anticancer drugs, we synthesized derivatives **7–24** possessing propargyl, 2-butynyl, 4-bromo-2-bu yl, and 4-hydroxy-2-butynyl groups with the aim to obtain more information about the influence of substituents on antiprolif ative activity in this class of compounds.

2

Results and discuss n

2.1

Chemistry

The synthesis of acetylenic thioquinolines **7**–**24** () was accomplished starting with 4-chloro-3-methylthio-quinoline **3** or 4-chloro-3-propargyl thioquinoline **4** or 4-chloro-3-(2-butynylthio)quinoline **5** and 4-chloro-3-(4-hydroxy-2-butynylthio)quinoline **6**.

Compounds **3–5** were prepared according to our previously reported methods. 4-Chloroquinoline **6** was synthesized as shown in . The starting **1** was prepared according to our published procedure. Treatment of **1** with sodium methoxide in DMSO at 20°C give sodium 4-chloro-3-quinolinethiolate **1-A** and 4-methoxy-3-methylthioquinoline **2**, which was removed by extraction. Sodium salt **1-A** after S-alkylation using 1-bromo-4-hydroxy-2-butyne gave with 65% yield **6**.

Compounds **3–6** were converted into **9–15** and **17–24** by nucleophilic displacement of chlorine atom by thiourea or selenourea in ethanol, hydrolysis of uronium salt **3-A** and subsequent S- or Se-alkylation of sodium salt **3-B** with methyl iodide or propargyl bromide or 1,4-dibromo-2-butyne or 1-bromo-4-hydroxy-2-butyne (). In the case of the synthesis of compounds **7**, **8**, and **16** the solution of sodium salt **3-B**, which was obtained from the reaction of corresponding **4** or **6** with thiourea or selenourea, was added to the ethanol solution containing an excess of 1,4-dibromo-2-butyne. Under these conditions, the concentration of **3-B** will always is low and it will favor the formation of corresponding 4-bromo-2-butynyl derivatives **7, 8** and **16**. The crude products were isolated from aqueous sodium hydroxide by filtration or extraction and separated by column chromatography.

2.2

Antiproliferative activity

ChemAxon compound

The eighteen compounds were tested in SRB or TT (in the case of leukemia cells) assay for their antiproliferative activity in vitro against two human cancer cell lines: SW70 colorectal adenocarcinoma), CCRF/CEM (leukemia) and two murine cancer cell lines: P388 (leukemia), B16 (melanoma). Th results of cytotoxic activity in vitro were expressed as an ID_{50} (μg/ml), that is, the concentration of compound, which inhibits th proliferation of 50% of tumor cells as compared to the control untreated cells. Cisplatin was applied as a referential cytotoxic a nt (positive test control). A value of less than 4μg/ml is considered as an antiproliferative activity criterion for synthetic co ounds. The results of the cytotoxicity studies are summarized in .

In general all the compounds obtained exhibited potent antiproliferative activity against human and murine cancer lines applied. 4-Chloro-3-(4-hydroxy-2-butynylthio)quinoline **6** exhibited high activity against P388 and CCRF/CEM and no antiproliferative activity against B16. As reported previously 4-chloro-3-(2-butynylthio)quinoline possessed lower cytotoxic activity than **6**. It indicated that 4-hydroxy-2-butynylthio group slightly influenced the activity than 2-butynylthio group. The

4

dictionary

Minimal confidence 0.1

- Query Term
- Chemistry
- Anatomy
- Cell Lines
- Diseases
- Potential IUPAC
- N2S ChemAxon
- N2S Opsin
- animals

Examples

Search terms

- “alkanes” PubMed **51,992**
- “alkanes” OCMiner **108,164**

- “heterocyclic” PubMed **24,711**
- “heterocyclic compounds” OCMiner **3,124,129**

- “spiro compounds” PubMed **33,528**
- “spiro compounds” OCMiner **157,396**

- “terpenes” PubMed **218,433**
- “terpenes” OCMiner **724,137**

Thanks For The Attention !

try at www.ocminer.com