Cheminformatics and Metabolism Resources

Janna Hastings
Group coordinator
Cheminformatics and Metabolism Team
Metabolomics

Measures occurrence and concentrations of many small molecules (metabolites) in an organism at once.
http://www.ebi.ac.uk/metabolights

MetaboLights
open-access, cross-species, cross-application

The EBI’s Metabolomics Database
Christoph Steinbeck, Jules Griffin
BBSRC BBR grant BB/I000933/1
MetaboLights study repository statistics

Data in MetaboLights

- Different organisms: 25
- Number of protocols: 346
- Number of samples: 10,510
- Private studies: 19
- Public studies: 32
- Total number of studies: 51

MetaboLights compounds references from other databases

- Total: 4,656
- ChEBI: 4,119
- GOLM: 53
- HMDB: 32
- KEGG: 16
- LIPID MAPS: 1
- Others: 29
- PubChem: 54
- Unknown: 188
- not mapped to any database: 164

Submitters

- Number of registered users: 216
MTBLC16010 - 5-oxoproline

DEFINITION
An oxoproline having the oxo group placed at the 5-position.
IUPAC NAME
5-oxoproline 5-oxopyrrolidine-2-carboxylic acid
CHEMICAL PROPERTIES
Chemical formula -
Chemical formula -
Average mass - 129.11400
SYNONYMS:
2-pyrrolidine-5-carboxylic acid 5-oxo-DL-proline 5-Oxoproline 5-OXOPROLINE 5-Pyrrolidone-2-carboxylic acid Glp Pyroglutamate Pyroglutamic acid
InChI=1S/C5H7NO3/c7-4-2-1-3(6-4)5(8)9/h3H,1-2H2,(H,6,7)(H,8,9)
Reference Layer
MTBLC16010 - 5-oxoproline

5-oxoproline - (CHEBI:16010)
Reference Layer
(+) artemisinin (CHEBI:223316)

ChEBI Name: (+)-artemisinin
ChEBI ID: CHEBI:223316
Definition: A sesquiterpene lactone obtained from sweet wormwood, Artemisia annua, which is used as an antimalarial for the treatment of multi-drug resistant strains of Plasmodium falciparum malaria.
Stars: ★★★ This entry has been manually annotated by the ChEBI Team.
Secondary ChEBI IDs: CHEBI:8718

Wikipedia

Artemisinin, also known as "Qinghaosu" (Chinese: 青蒿素), and its derivatives are a group of drugs that possess the most rapid action of all current drugs against Plasmodium falciparum malaria. Treatments containing an artemisinin derivative (artemisinin-combination therapies, ACTs) are now standard treatment worldwide for P. falciparum malaria. The starting compound artemisinin is isolated from the plant Artemisia annua, an herb described in Chinese traditional medicine. Chemically, artemisinin is a sesquiterpene lactone containing an unusual peroxide bridge. This peroxide is believed to be responsible for the drug's mechanism of action. Few other natural compounds with such a peroxide bridge are known. (Ascaridole is another.) Use of the drug by itself as a monotherapy is explicitly discouraged by the World Health Organization, as there have been signs that malarial parasites are developing resistance to the drug. Therapies that combine artemisinin with some other antimalarial drug are the preferred treatment for malaria and are both effective and well tolerated in patients. The drug is also increasingly being used in Plasmodium vivax malaria, as well as being a topic of research in cancer treatment.

Read full article at Wikipedia
What does ChEBI provide?

Names and synonyms
- caffeine
- 1,3,7-trimethylxanthine
- methyltheobromine

Chemical data
- Formula: C8H10N4O2
- Charge: 0
- Mass: 194.19

Chemical Informatics
- InChI=1/C8H10N4O2/c1-10-4-9-6-5(10)7(13)12(3)8(14)11(6)2/h4H,1-3H3
- SMILES CN1C(=O)N(C)c2ncn(C)c2C1=O

Ontology – classifications
- metabolite
- CNS stimulant
- trimethylxanthines

Links to more information in other databases
- MSDchem: CFF
- KEGG DRUG: D00528

Chemical structures and visualisations
Increasing focus on natural products

Avicularin (CHEBI:65460)

- **CHEBI Name**: Avicularin
- **CHEBI ID**: CHEBI:65460
- **Definition**: A natural product found particularly in *Juglans regia* and *Foeniculum vulgare*.
- **Stars**: This entity has been manually annotated by a third party.
- **Secondary CHEBI IDs**: CHEBI:545755

**Formula**: C20H18O11

- **Net Charge**: 0
- **Average Mass**: 434.35030

**InChI**: InChI=1S/C20H18O11/c21-6-13-15(26)17(28)20(30-13)31-19-16(27)14-11(25)4-8(22)6-12(14)20-18(19)7-1-2-9(23)10(24)3-7/h1-5,13,16,17,20-26,28H,6H2/t13-,15-,17+,20-/m0/s1

**InChIKey**: BDCDNTVZSILEOY-UXYN5RIQZSA-N

**SMILES**: OC[C@H]1O[C@H]((Oc2c(oc3cc(O)cc(O)c3c2=O)-c2cc(O)c(O)c2)[C@H]((O)[C@H]1O
Global Collaboration in Metabolomics and the BioSciences
COSMOS
COOrdination of Standards in MetabolOmicS

• 2 mio Euros for meetings and coordination
• 90.000 Euros for travel for global stakeholders

INFRA-2012-3.3. "Coordination actions, conferences and studies supporting policy development, including international cooperation, for e-Infrastructures".
Research
PAMELA pipeline for metabolome inference

DB integration
Text Mining
Enumeration
Validation

Computer-Assisted Structure Elucidation (CASE)

The Chemistry Development Kit (CDK): An Open Source Java Library for Structural Cheminformatics

Input/Output
- I/O (CML, MDL Moffile, SDF, PDB)
- SMILES
- InChI

Modelling
- 3D Model-Builder
- Atom-Typing
- Force-Field
- Representation of Biomolecular Structures

Visualization
- Structure-Diagram-Layout (SDG)
- 2D Rendering
- 3D Rendering

Library Enumeration
- Deterministic Isomer generator
- Stochastic Structure Generators via:
  - Simulated Annealing
  - Genetic Algorithms

Chemical Graphs
- Isomorphism detection
- Maximum-Common-Substructure Searches
- SMARTS- and Substructure searches
- Ring searches
- Aromaticity detection

Properties
- Fingerprinting
- > 70 QSAR-Descriptors
- QSAR model building

http://cdk.sourceforge.net
Feature Extraction and Spectrum Processing

Stochastic Searching for Structure Elucidation

Fitness Evaluation (Scoring)

\[ S_{\text{total}} = S_{\text{NMR-HMBC}} + S_{\text{NMR-HH COSY}} + \\
S_{\text{NMR-Shift}} + S_{\text{Symmetry}} + S_{\text{MassSpec}} + \\
S_{\text{Features}} \]

Polycarpol (C\textsubscript{30}H\textsubscript{48}O\textsubscript{2}).

• **Natural Product-likeness classification** and integrated it into Taverna workflow tool
• Included in second version of SENECA CASE

Desktop Application

- Simplify **editing** of genome-scale metabolic model
- Backed by the CDK providing **structure** representation
- Database free, access to common **resources** resolved automatically through web and local instances
- Export to **annotated** SBML
- Use the structure to rapidly **merge, compare and complete** models (wip)

Thank You