



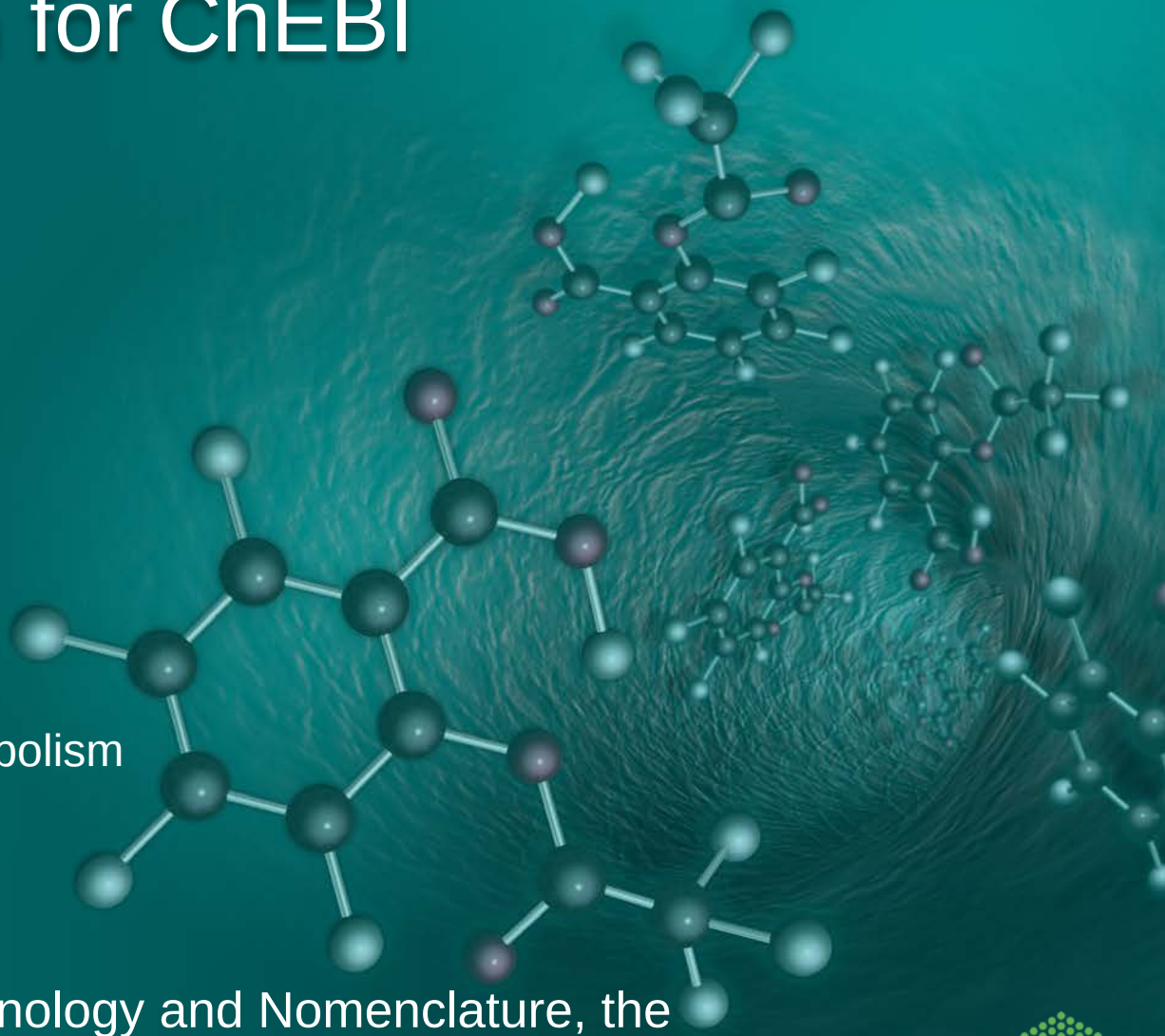
Defining chemical classes in OWL-based English for ChEBI

Janna Hastings

Coordinator,
Cheminformatics and Metabolism

www.ebi.ac.uk/chebi

What's in a Name: Terminology and Nomenclature, the
unsung heroes of open innovation, 21 October 2014



Chemical Entities of Biological Interest

EMBL-EBI

Services Research Training About us

ChEBI

The database and ontology of Chemical Entities of Biological Interest

EBI > Databases > Small Molecules > ChEBI > Home

ChEBI Home

Definitions, relationships, hierarchy

E.g. metabolites, drugs, pesticides

Freely available online, available for download in full

Search for ★★★★★ only: ☐ All in ChEBI ☒ ChEBI and Partner Databases ☐

Search ChEBI

Wildcard character is *
Example: "iron*", "InChI=1S/H2O/h1H2"

[Advanced Search](#) | [About ChEBI](#)

Chemical Entities of Biological Interest (ChEBI) is a freely available dictionary of molecular entities focused on 'small' chemical compounds.

Low molecular weight, i.e. no proteins

38,215 entries last release

Entity of the month
3rd March 2014
5-Hydroxymethylfurfural

5-Hydroxymethylfurfural (CHEBI:412516), commonly abbreviated to 5-HMF, was one of the

03 March 2014 - ChEBI release 113
ChEBI release 113 is now available, with 38215 fully annotated entities.
See our

What does ChEBI provide?

Names and synonyms

caffeine
1,3,7-trimethylxanthine
methyltheobromine

Ontology – classifications

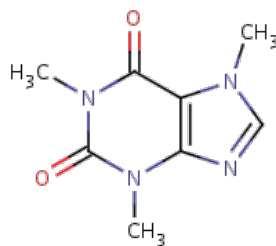
metabolite
CNS stimulant
trimethylxanthines

Chemical data

Formula: C₈H₁₀N₄O₂
Charge: 0
Mass: 194.19

Links to more information

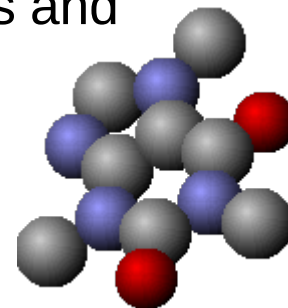
MSDchem: CFF
KEGG DRUG: D00528
PubMed citations



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

Chemical structures and visualisations



Example ChEBI entry page

EBI > Databases > Small Molecules > ChEBI > Main

caffeine (CHEBI:27732)

Search ChEBI here!

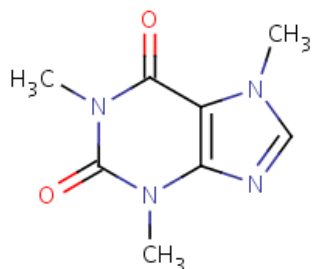
Search ChEBI

Search for ☆☆☆ only ☐ All in ChEBI ☒

Main

ChEBI Ontology

Automatic Xrefs



ChEBI Name [?](#) **caffeine**

ChEBI ID [?](#) **CHEBI:27732**

Definition [?](#) A trimethylxanthine in which the three methyl groups are located at positions 1, 3, and 7.

Stars [?](#) ☆☆☆ This entity has been manually annotated by the ChEBI Team.

Secondary ChEBI IDs [?](#) CHEBI:41472, CHEBI:116485, CHEBI:22982, CHEBI:3295

Supplier Information  [eMolecules:27517656](#), [eMolecules:493944](#), [ZINC00001084](#)

See structure as: ☒ Image ☐ Applet

 [Download Molfile](#)

- [Find compounds which contain this structure](#)
- [Find compounds which resemble this structure](#)

[more structures >>](#)

[Wikipedia](#)

[License](#)

Caffeine is a bitter, white [crystalline xanthine alkaloid](#) and a [stimulant](#) drug. Caffeine is found in varying quantities in the seeds, leaves, and fruit of some plants, where it acts as a natural pesticide that paralyzes and kills certain insects feeding on the plants, as well as enhancing the reward memory of pollinators. It is most commonly consumed by humans in infusions extracted from the seed of the [coffee plant](#) and the leaves of the [tea bush](#), as well as from various foods and drinks containing products derived from the [kola nut](#). Other sources include [yerba maté](#), [guarana](#) berries, [guayusa](#), and the [yaupon holly](#). In humans, caffeine acts as a [central nervous system](#) stimulant, temporarily warding off drowsiness and restoring alertness. Studies have shown that caffeine potentially induces [chromosomal aberrations](#), and shows both [teratogenic](#) and [mutagenic](#) properties. It is the world's most widely consumed [psychoactive drug](#), but unlike many other psychoactive substances, it is legal and unregulated in nearly all parts of the world. Beverages containing caffeine, such as [coffee](#), [tea](#), [soft drinks](#), and [energy drinks](#), enjoy great popularity. In North America, 90% of adults consume caffeine daily. Part of the reason caffeine is classified by the [Food and Drug Administration](#) as [GRAS](#) (generally recognized as safe) is that toxic doses (over 10 grams for an average adult) are much higher than typically used doses (less than 500 milligrams). Ordinary consumption can have low health risks, even when carried on for years – there may be a modest protective effect against some diseases, including [Parkinson's disease](#) and certain types of [cancer](#). The [American Congress of Obstetricians and Gynecologists](#) (ACOG) concluded in 2010 that caffeine consumption is safe up to 200 mg per day in pregnant women. Caffeine has [pressor](#) and mild [diuretic](#) effects when administered to people who are not used to it, but regular users develop a tolerance to this effect, and studies have generally failed to support the common notion that ordinary

Example entry page (continued)

Formula ?	Source
C8H10N4O2	KEGG COMPOUND
Net Charge ?	0
Average Mass ?	194.19076
InChI ?	InChI=1S/C8H10N4O2/c1-10-4-9-6-5(10)7(13)12(3)8(14)11(6)2/h4H,1-3H3
InChIKey ?	RYVVLZVUVIJVGH-UHFFFAOYSA-N
SMILES ?	<chem>Cn1cnc2n(C)c(=O)n(C)c(=O)c12</chem>
Roles Classification ?	
Chemical Role(s):	EC 3.1.4.* (phosphoric diester hydrolase) inhibitor An EC 3.1.* (ester hydrolase) inhibitor that interferes with the action of a phosphoric diester hydrolase (EC 3.1.4.*). EC 2.7.11.1 (non-specific serine/threonine protein kinase) inhibitor An EC 2.7.11.* (protein-serine/threonine kinase) inhibitor that interferes with the action of non-specific serine/threonine protein kinase (EC 2.7.11.1), a kinase enzyme involved in phosphorylation of hydroxy group of serine or threonine.
Biological Role(s):	adenosine receptor antagonist An antagonist at any adenosine receptor. ryanodine receptor modulator EC 3.1.4.* (phosphoric diester hydrolase) inhibitor An EC 3.1.* (ester hydrolase) inhibitor that interferes with the action of a phosphoric diester hydrolase (EC 3.1.4.*). EC 2.7.11.1 (non-specific serine/threonine protein kinase) inhibitor An EC 2.7.11.* (protein-serine/threonine kinase) inhibitor that interferes with the action of non-specific serine/threonine protein kinase (EC 2.7.11.1), a kinase enzyme involved in phosphorylation of hydroxy group of serine or threonine. ryanodine receptor agonist A ryanodine receptor modulator which activates the receptor. Ryanodine receptors (RyRs) act as selective ion channels, modulating the release of calcium. Activating the receptors causes the release of calcium, so depleting internal calcium and ultimately preventing further muscle contraction. fungal secondary metabolite Any fungal metabolite that is not directly involved in the normal growth, development or reproduction of fungi. adenosine A2A receptor antagonist An antagonist at the A2A receptor.

Example entry page (continued)

IUPAC Name

1,3,7-trimethyl-3,7-dihydro-1H-purine-2,6-dione

Synonyms

1,3,7-trimethyl-2,6-dioxopurine

1,3,7-trimethylpurine-2,6-dione

1,3,7-trimethylxanthine

1-methyltheobromine

3,7-Dihydro-1,3,7-trimethyl-1H-purin-2,6-dion 

7-methyltheophylline

Sources

ChemIDplus 

IUPHAR

NIST Chemistry WebBook 

ChemIDplus 

NIST Chemistry WebBook 

NIST Chemistry WebBook 

Database Links

[1-3-7-TRIMETHYLXANTHINE](#)

[116485](#)

[C07481](#)

[Caffeine](#)

[CFF](#)

[D00528](#)

[DB00201](#)

[HMDB01847](#)

[View more database links](#)

Databases

MetaCyc

ChEMBL

KEGG COMPOUND

Wikipedia

PDBeChem

KEGG DRUG

DrugBank

HMDB

Registry Numbers

103040

Types

Gmelin Registry Number

Sources

Gmelin

17705

Reaxys Registry Number

Reaxys

17705

Beilstein Registry Number

Beilstein

58-08-2

CAS Registry Number

KEGG COMPOUND

[58-08-2](#)

CAS Registry Number

ChemIDplus

[58-08-2](#)

CAS Registry Number

NIST Chemistry WebBook

Citations

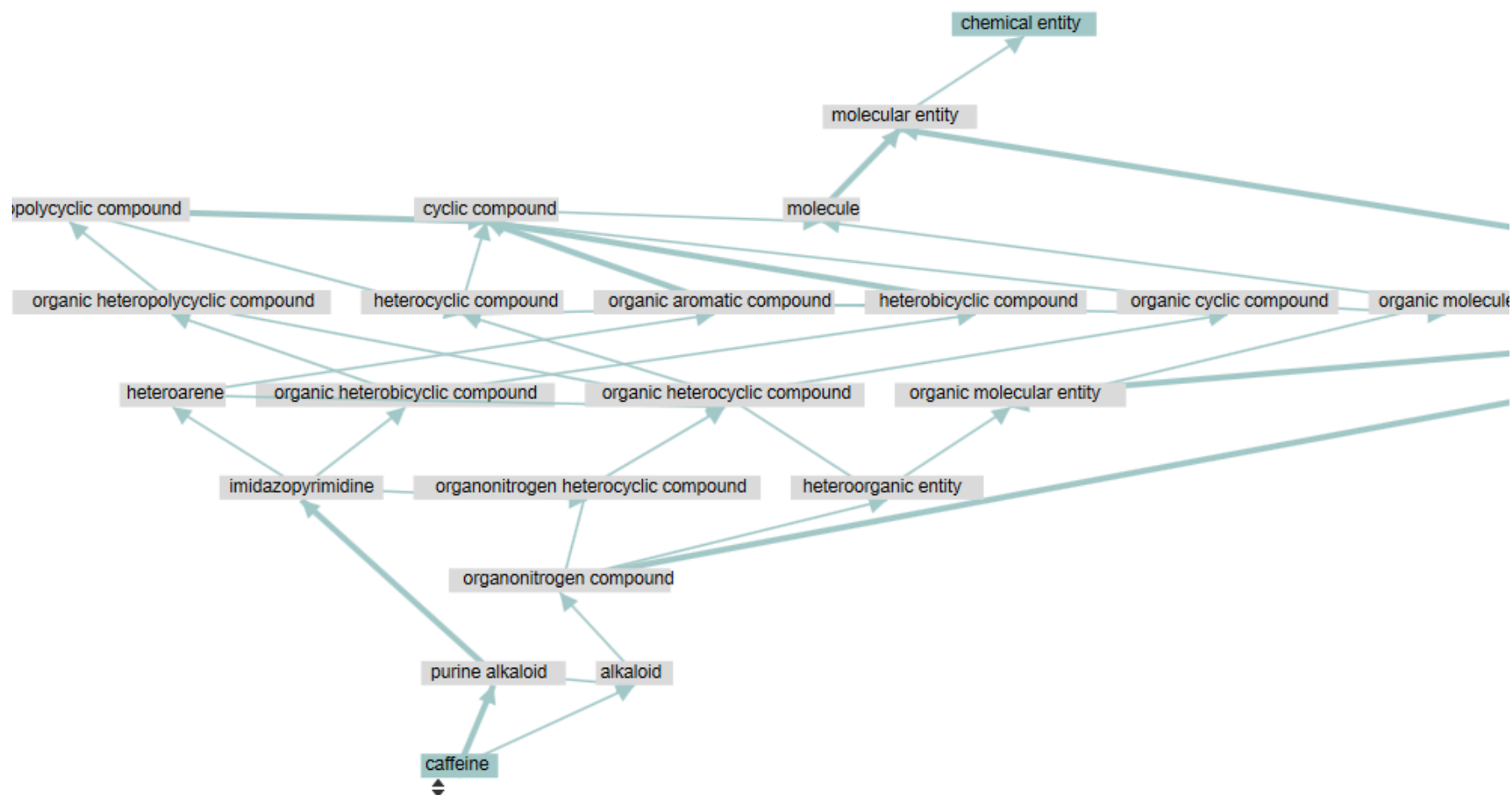
Bara AI, Barley EA (2000)

Caffeine for asthma.

The Cochrane database of systematic reviews CD001112 [[PubMed:10796597](#)]

[\[show Abstract\]](#)

Structure-based chemical classes in ChEBI



The Web Ontology Language (OWL)

chebi (ftp://ftp.ebi.ac.uk/pub/databases/chebi/ontology/chebi.owl) : [ftp://ftp.ebi.ac.uk/pub/databases/chebi/ontology/chebi.owl]

File Edit View Reasoner Tools Refactor Window Help

chebi (ftp://ftp.ebi.ac.uk/pub/databases/chebi/ontology/chebi.owl) caffeine

Active Ontology Entities Classes Object Properties Data Properties Annotation Properties Individuals OWLViz DL Query OntoGraf

Class hierarchy Class hierarchy (inferred) General class axioms

Class hierarchy: caffeine

- 'purine alkaloid'
- 7-methylxanthine
- caffeine
- 'lupinic acid'
- methylxanthine
- 'purine N-oxides'
- 'purine nucleobase'
- 'purine nucleoside'
- 'purine nucleoside bisphosphate'
- 'purine nucleotide'
- purine-6-carboxamide
- purmorphamine

Selected entity Rules

Class Annotations Class Usage

Annotations: caffeine

Annotations +

- label [type: string] caffeine
- altid [type: string] CHEBI:22982
- altid [type: string] CHEBI:3295
- altid [type: string]

Description: caffeine

Equivalent To +

SubClass Of +

- 'has role' some 'adenosine A2A receptor antagonist'
- 'has role' some 'adenosine receptor antagonist'
- 'has role' some 'central nervous system stimulant'
- 'has role' some 'EC 2.7.11.1 (non-specific serine/threonine protein kinase) inhibitor'
- 'has role' some 'EC 3.1.4.* (phosphoric diester hydrolase) inhibitor'
- 'has role' some 'food additive'
- 'has role' some 'fungal metabolite'

Individuals by type Annotation property hierarchy Datatypes

Object property hierarchy Data property hierarchy

Object property hierarchy: 'has role'

topObjectProperty

- 'has functional parent'
- 'has part'
- 'has parent hydride'
- 'has role'
- 'is conjugate acid of'
- 'is conjugate base of'

Classes

Metadata

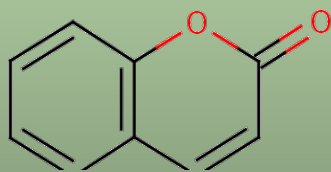
Definitions

Relationships

Axioms (these are
quantified
relationships)

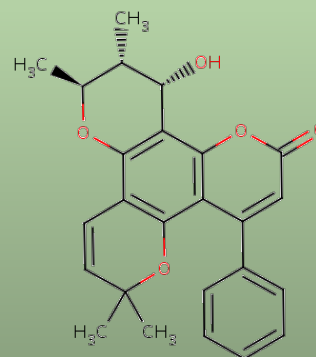
The logic of structural chemical definitions

coumarin



has part

soulattrolide



is a

coumarins

Gen Pharmacol. 1996 Jun;27(4):713-22.

Pharmacological and biochemical actions of simple coumarins: natural products with therapeutic potential.

Hoult JR¹, Payá M.

⊕ Author information

Abstract

1. More than 300 coumarins have been identified from natural sources, especially green plants. The pharmacological and biochemical properties and therapeutic applications of simple coumarins depend upon the pattern of substitution. More complex related compounds based on the coumarin nucleus include the dicoumarol/warfarin anticoagulants, aflatoxins and the psoralens (photosensitizing agents). 2. Coumarin itself (1,2-benzopyrone) has long-established efficacy in slow-onset long-term reduction of lymphoedema in man, as confirmed in recent double-blind trials against elephantiasis and postmastectomy swelling of the arm. The mechanism of action is uncertain, but may involve macrophage-induced proteolysis of

A language for definitions: names mapped to structures in ChEBI

thiadiazoles:

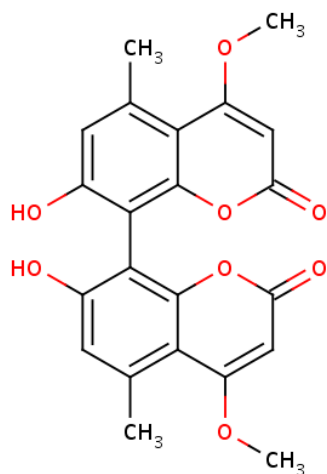
molecular_entity *and* has_part
some (1,2,3-thiadiazole *or* 1,2,4-thiadiazole
or 1,2,5-thiadiazole *or* 1,3,4-thiadiazole)

bicoumarin: organic_molecular_entity *and*
has_part *exactly* 2 coumarin

Parts

Counts (min, max
and exact)

Logical operators



orlandin

A language for definitions (2)

organic ion: organic_molecular_entity and
(has_charge some int[>0] or has_charge some int[<0])

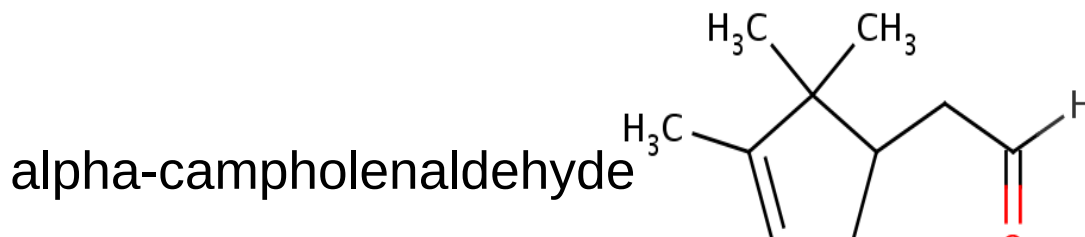
Value ranges

Properties: charges, cycles

monocyclic compound: molecular_entity and
has_cycles value "1"^^int

XSD datatypes

Exact values



Capturing definitions / axioms

Type a name or ID:

Select

metala

metalaxyl

metalaxyl-M

or select a class:

acyl-CoA
alkaloids
agonist
amino_oligosaccharide
antagonist
anti-HIV_agent
anti-inflammatory_agent
antibacterial_agent
antibiotic
antineoplastic_agent
antioxidant
carboxylic_acid
coumarins
fatty_acid
flavonoids
ganglioside
enzyme_activator
enzyme_inhibitor
organophosphate_oxoxani
phosphoglycerolipid
phosphosphingolipid
secondary_metabolite
sesquiterpenoids
steroid
triterpenoids

Add axiom:

Add



☐ Perform fuzzy search for suggestion list.

metalaxyl

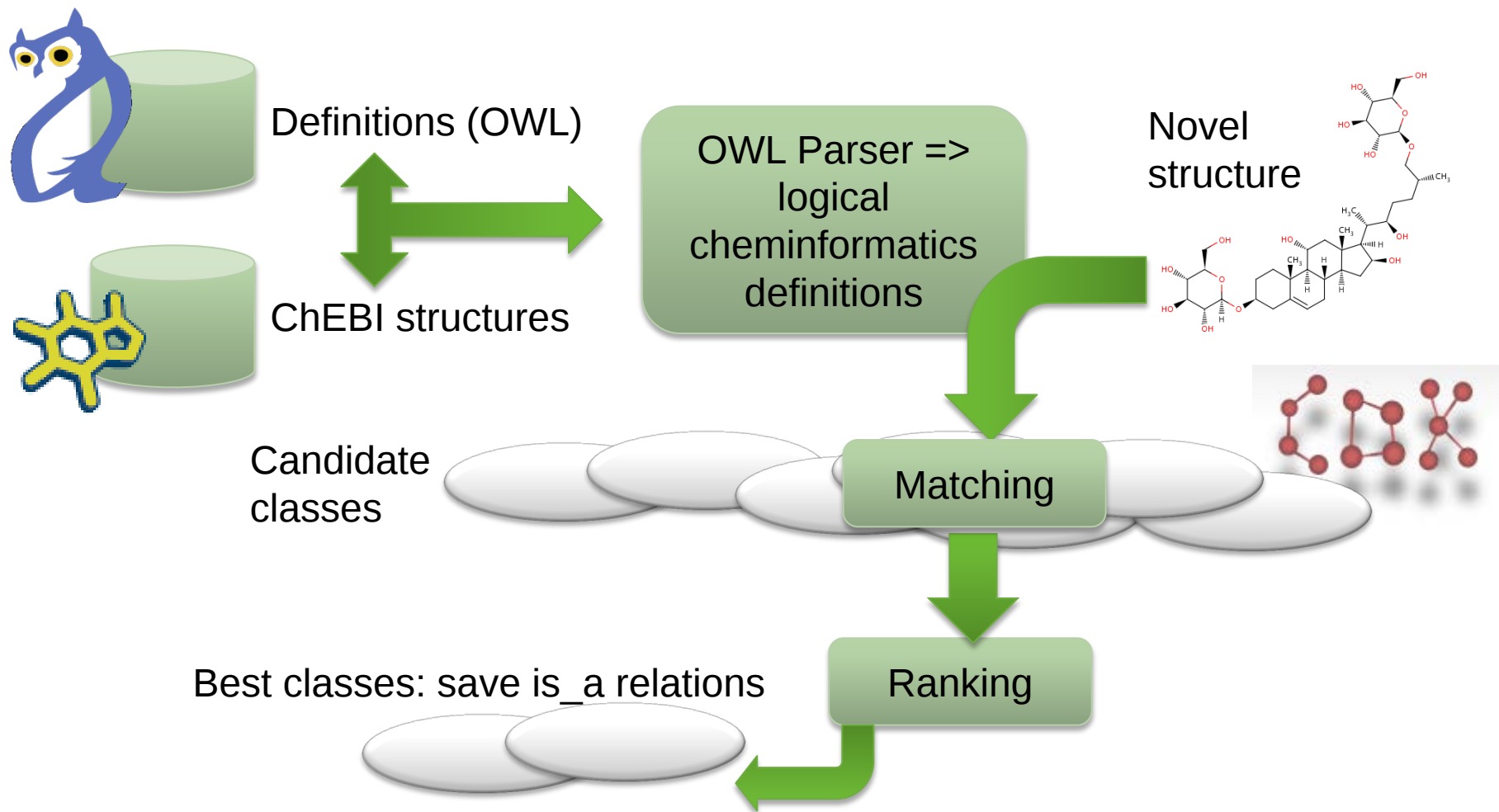
SubClass Of Axioms

- primary_amide
- amino_acid_ester
- has_role some agrochemical
- has_role some amide_fungicide

base or aci

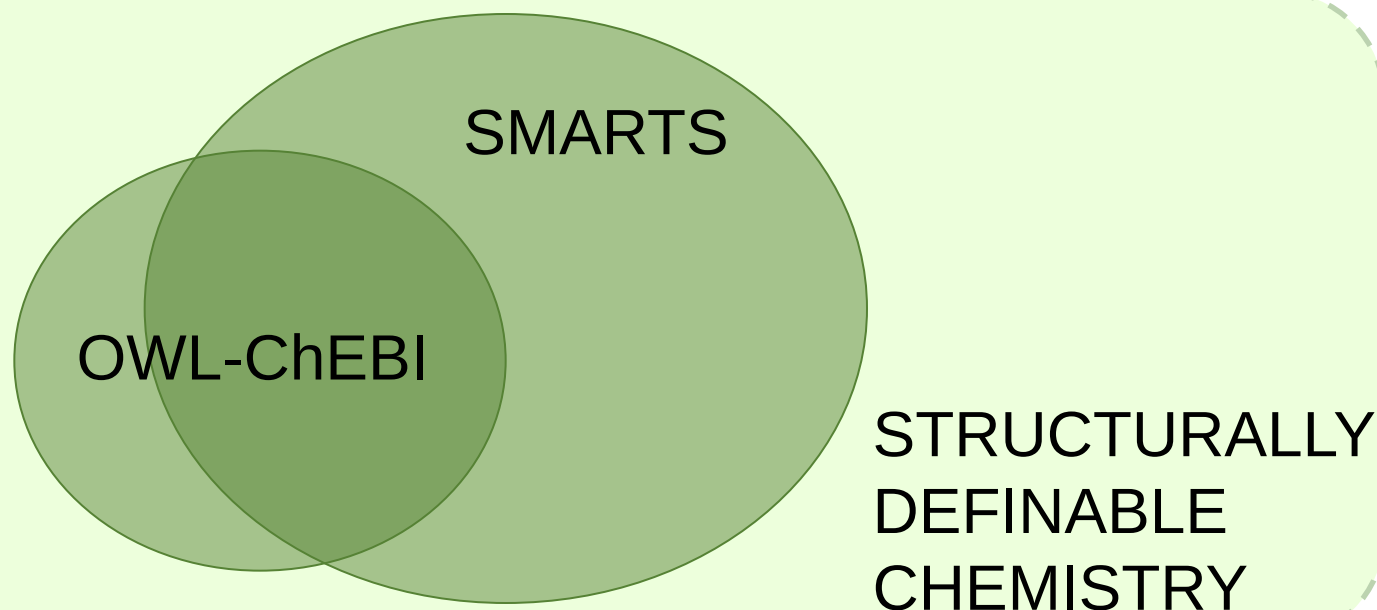
acid
acitretin
acid_anhydride
acidity_regulator
acid-base_indicator
bacitracin
bacillithiol
bacitracin_A
pacific_blue
bacillosamine
acidic_glycosphingolipid
Pacificin_L(rel)

Interpreting OWL-English definitions



An extensible, human-readable language

- “nouns” are classes in ChEBI
- “verbs” are interpretable relationships interpretable
- Grammar rules are defined by the OWL language



Limitations – No support for:

- OWL allows at most binary relations.
- Unable to differentiate specific attachment points
- Relative locations
- Stereochemistry (unless explicit in substructure)
- “Fuzzy” substructures
- Polymers

Acknowledgements – Thanks!

ChEBI team:

Christoph Steinbeck
Gareth Owen
Adriano Dekker
Namrata Kale
Steve Turner
Venkatesh Muthukrishnan

Collaborators:

Colin Batchelor, RSC
Lian Duan, ETH
Leonid Chepelev, Ottawa
Michel Dumontier, Stanford
Despoina Magka, Oxford
Ilinca Tudose and John May, EBI

Funding:

BBSRC “Continued
development of ChEBI towards
better usability for the systems
biology and metabolic
modelling communities”
BB/K019783/1



Thank you for listening!

Questions?

chebi-help@ebi.ac.uk

hastings@ebi.ac.uk

