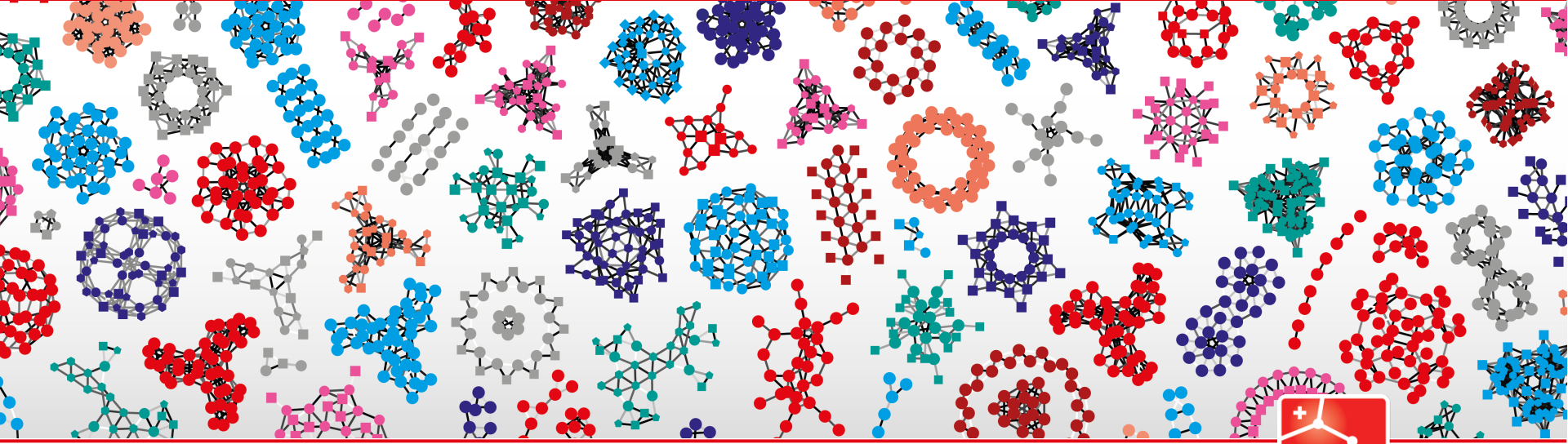


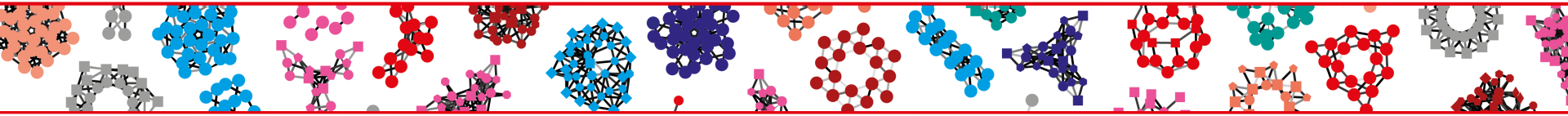
Swiss Institute of Bioinformatics

SPARQLing Rhea

Anne Morgat, Swiss-Prot group

elixir SWITZERLAND | swissprot | Rhea

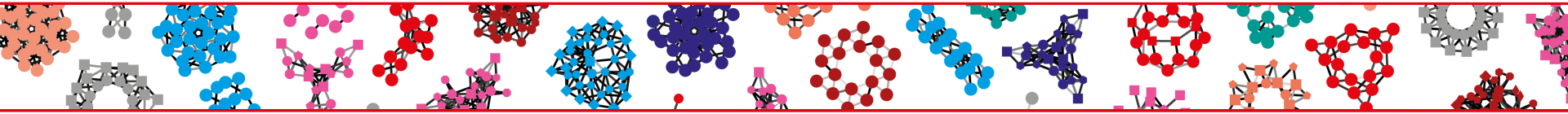
www.sib.swiss



Accompanying Jupyter notebook:

https://github.com/sib-swiss/sparql-training/tree/master/rhea/SWAT4HCLS_2019

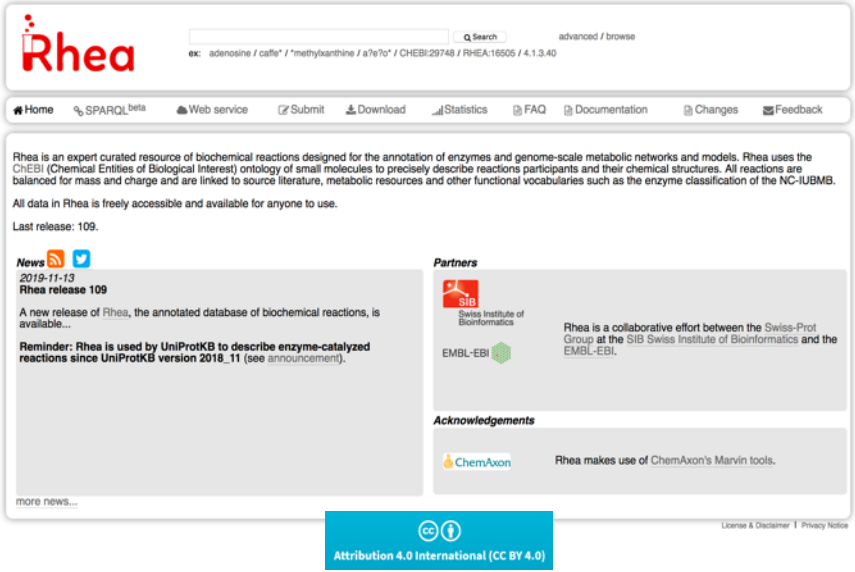
Overview



Rhea, a biochemical reactions knowledgebase

Rhea is an **expert curated knowledgebase of chemical and transport reactions of biological interest** that can be used for enzyme annotation, genome-scale metabolic modeling and omics-related analysis

- **12,290 biochemical reactions**
- **10,795 compounds**
- **14,082** curated **literature references**
- **covers all organisms**
- Used by **UniProtKB**, **ChEBI, SwissLipids, EBI Enzyme Portal, MetaboLights, MetaNetX,...** *and soon by GO (Gene Ontology)*



https://www.rhea-db.org

Rhea reactions: balanced – non redundant

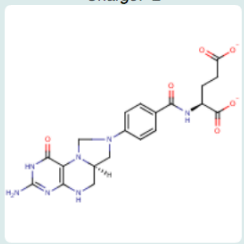
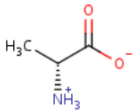
RHEA:10064 (APPROVED)

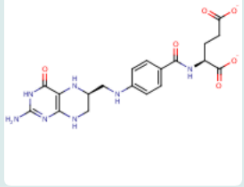
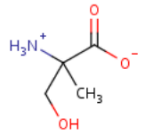
(6R)-5,10-methylene-5,6,7,8-tetrahydrofolate + D-alanine + H2O = (6S)-5,6,7,8-tetrahydrofolate + 2-methylserine

Last modified: 2017-10-18. **Chemically balanced:** yes.

(6*R*)-5,10-methylene-5,6,7,8-tetrahydrofolate + **D-alanine** + **H_2O** = **(6*S*)-5,6,7,8-tetrahydrofolate** + **2-methylserine**

(6*R*)-5,10-methylene-5,6,7,8-tetrahydrofolate	D-alanine	H_2O	(6*S*)-5,6,7,8-tetrahydrofolate	2-methylserine
CHEBI:15636	CHEBI:57416	CHEBI:15377	CHEBI:57453	CHEBI:58275
Formula: C20H21N7O6	*Formula:* C3H7NO2	*Formula:* H2O	*Formula:* C19H21N7O6	*Formula:* C4H9NO3
Charge: -2	*Charge:* 0	*Charge:* 0	*Charge:* -2	*Charge:* 0

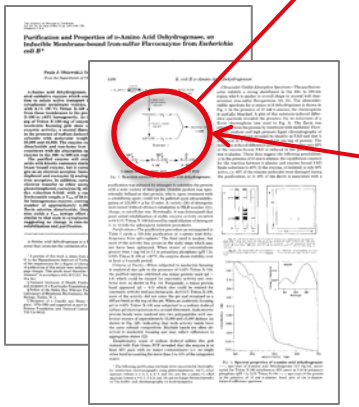



formula: $C_{23}\ H_{30}\ N_8\ O_9$, charge: -2

formula: $C_{23}\ H_{30}\ N_8\ O_9$, charge: -2

Rhea curation: describing chemical reactions using ChEBI

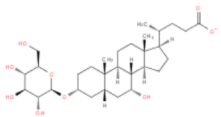
5,10-methylenetetrahydrofolate + D-alanine + H_2O = tetrahydrofolate + 2-methylserine



Rhea curators

www.ebi.ac.uk/chebi

CHEBI:142610 - chenodeoxycholate-3-*O*-β-D-glucoside(1−)



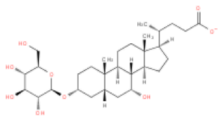
ChEBI Name	**chenodeoxycholate-3-*O*-β-D-glucoside(1−)**
ChEBI ID	**CHEBI:142610**
ChEBI ASCII Name	chenodeoxycholate-3-O-beta-D-glucoside(1-)
Definition	A β-D-glucoside resulting from the formal condensation of the 1*R*-hydroxy group of β-D-glucose with the 3*R*-hydroxy group of chenodeoxycholate.
Stars	⭐⭐⭐ This entity has been manually annotated by the ChEBI Team.
Submitter	Lucila Aimo
Supplier Information	No supplier information found for this compound.
Download	Molfile XML SDF

- Find compounds which contain this structure
- Find compounds which resemble this structure
- Take structure to the Advanced Search

Formula	C30H49O9
Net Charge	-1
Average Mass	553.706
Monoisotopic Mass	553.33821
InChI	InChI=1S/C30H50O9/c1-15(4-7-23(33)34)18-5-6-19-24-20(9-11-30(18,19)3)29(2)10-8-17(12-16(29)13-21(24)32)38-28-27(37)26(36)25(35)22(14-31)39-28/h15-22,24-28,31-32,35-37H,4-14H2,1-3H3,(H,33,34)/p-1/t15-,16+,17-,18-,19+,20+,21-,22-,24+,25-,26+,27-,28-,29+,30-/m1/s1
InChIKey	QRLIJDGVRXVHQD-UVMPBBQUSA-M
SMILES	C1[C@@]2([C@]3(CC[C@]4([C@]([C@@]3([C@@H](C[C@@]2(C[C@@H](C1)O[C@H]5[C@@H]([C@H]([C@@H]([C@H](O5)CO)O)O)O)[H])O)[H])(CC[C@@]4([C@@H](CCC([O-])=O)C)[H])[H])C)[H])C

CHEBI:142610 - chenodeoxycholate-3-*O*-β-D-glucoside(1−)

Main | ChEBI Ontology | Automatic Xrefs | Reactions | Pathways | Models



ChEBI Name	**chenodeoxycholate-3-*O*-β-D-glucoside(1−)**
ChEBI ID	**CHEBI:142610**
ChEBI ASCII Name	chenodeoxycholate-3-O-beta-D-glucoside(1-)
Definition	A β-D-glucoside resulting from the formal condensation of the 1*R*-hydroxy group of β-D-glucose with the 3*R*-hydroxy group of chenodeoxycholate.
Stars	⭐⭐⭐ This entity has been manually annotated by the ChEBI Team.
Submitter	Lucila Aimo
Supplier Information	No supplier information found for this compound.
Download	Molfile XML SDF

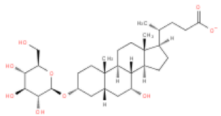
- Find compounds which contain this structure
- Find compounds which resemble this structure
- Take structure to the Advanced Search

Formula	C30H49O9
Net Charge	-1
Average Mass	553.706
Monoisotopic Mass	553.33821
InChI	InChI=1S/C30H50O9/c1-15(4-7-23(33)34)18-5-6-19-24-20(9-11-30(18,19)3)29(2)10-8-17(12-16(29)13-21(24)32)38-28-27(37)26(36)25(35)22(14-31)39-28/h15-22,24-28,31-32,35-37H,4-14H2,1-3H3,(H,33,34)/p-1/t15-,16+,17-,18-,19+,20+,21-,22-,24+,25-,26+,27-,28-,29+,30-/m1/s1
InChIKey	QRLIJDGVRXVHQD-UVMPBBQUSA-M
SMILES	C1[C@@]2([C@]3(CC[C@]4([C@]([C@@]3([C@@H](C[C@@]2(C[C@@H](C1)O[C@H]5[C@@H]([C@H]([C@@H]([C@H](O5)CO)O)O)O)[H])O)[H])(CC[C@@]4([C@@H](CCC([O-])=O)C)[H])[H])C)[H])C

ChEBI:142610 - relationships

CHEBI:142610 - chenodeoxycholate-3-*O*-β-D-glucoside(1−)

Main | ChEBI Ontology | Automatic Xrefs | Reactions | Pathways | Models



ChEBI Name	**chenodeoxycholate-3-*O*-β-D-glucoside(1−)**
ChEBI ID	**CHEBI:142610**
ChEBI ASCII Name	chenodeoxycholate-3-O-beta-D-glucoside(1-)
Definition	A β-D-glucoside resulting from the formal condensation of the 1*R*-hydroxy group of β-D-glucose with the 3*R*-hydroxy group of chenodeoxycholate.
Stars	⭐⭐⭐ This entity has been manually annotated by the ChEBI Team.
Submitter	Lucila Aimo
Supplier Information	No supplier information found for this compound.
Download	Molfile XML SDF

- Find compounds which contain this structure
- Find compounds which resemble this structure
- Take structure to the Advanced Search

Formula	C30H49O9
Net Charge	-1
Average Mass	553.706
Monoisotopic Mass	553.33821
InChI	InChI=1S/C30H50O9/c1-15(4-7-23(33)34)18-5-6-19-24-20(9-11-30(18,19)3)29(2)10-8-17(12-16(29)13-21(24)32)38-28-27(37)26(36)25(35)22(14-31)39-28/h15-22,24-28,31-32,35-37H,4-14H2,1-3H3,(H,33,34)/p-1/t15-,16+,17-,18-,19+,20+,21-,22-,24+,25-,26+,27-,28-,29+,30-/m1/s1
InChIKey	QRLIJDGVRXVHQD-UVMPBBQUSA-M
SMILES	C1[C@@]2([C@]3(CC[C@]4([C@]([C@@]3([C@@H](C[C@@]2(C[C@@H](C1)O[C@H]5[C@@H]([C@H]([C@@H]([C@H](O5)CO)O)O)O)[H])O)[H])(CC[C@@]4([C@@H](CCC([O-])=O)C)[H])[H])C)[H])C

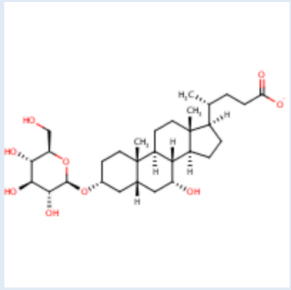
CHEBI:142610 - chenodeoxycholate-3-*O*-β-D-glucoside(1−)

Main | ChEBI Ontology | Automatic Xrefs | **Reactions** | Pathways | Models

select reaction: beta-D-glucosyl-(1->3)-O-chenodeoxycholate + H(2)O = chenodeoxycholate + D-glucose

RHEA:58340

beta-D-glucosyl-(1->3)-O-chenodeoxycholate
CHEBI:142610
Formula: C30H49O9
Charge: -1

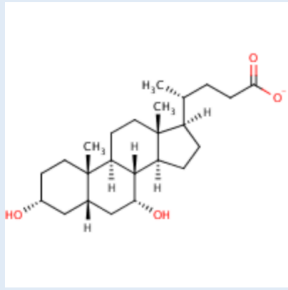


+

H2O
CHEBI:15377
Formula: H2O
Charge: 0

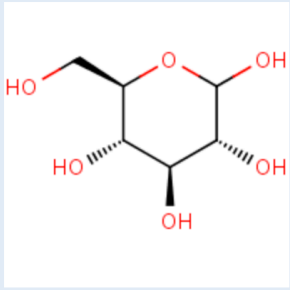
=

chenodeoxycholate
CHEBI:36234
Formula: C24H39O4
Charge: -1



+

D-glucose
CHEBI:4167
Formula: C6H12O6
Charge: 0



CHEBI:142610 - chenodeoxycholate-3-*O*-β-D-glucoside(1−)

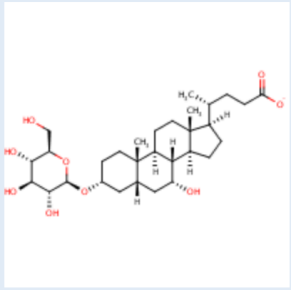
Main | ChEBI Ontology | Automatic Xrefs | **Reactions** | Pathways | Models

select reaction:

- ✓ beta-D-glucosyl-(1->3)-O-chenodeoxycholate + H(2)O = chenodeoxycholate + D-glucose
- chenodeoxycholate + dolichyl beta-D-glucosyl phosphate = beta-D-glucosyl-(1->3)-O-chenodeoxycholate + dolichyl phosphate + H(+)
- chenodeoxycholate + octyl beta-D-glucose = beta-D-glucosyl-(1->3)-O-chenodeoxycholate + octan-1-ol

RHEA:583

beta-D-glucosyl-(1->3)-O-chenodeoxycholate
CHEBI:142610
Formula: C30H49O9
Charge: -1

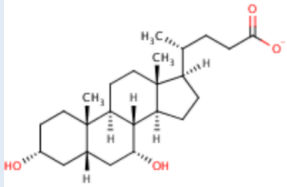


+

H2O
CHEBI:15377
Formula: H2O
Charge: 0

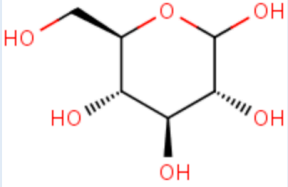
=

chenodeoxycholate
CHEBI:36234
Formula: C24H39O4
Charge: -1



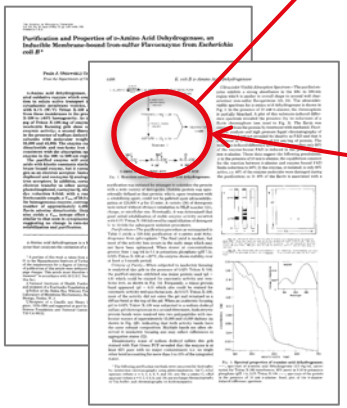
+

D-glucose
CHEBI:4167
Formula: C6H12O6
Charge: 0



Rhea curation: choice of reaction participants

5,10-methylenetetrahydrofolate + D-alanine + H_2O = tetrahydrofolate + 2-methylserine



Rhea curators

D-alanine

D-alaninate CHEBI:32435

CHEBI:32436 D-alaninium

ChEBI

D-alanine CHEBI:15570

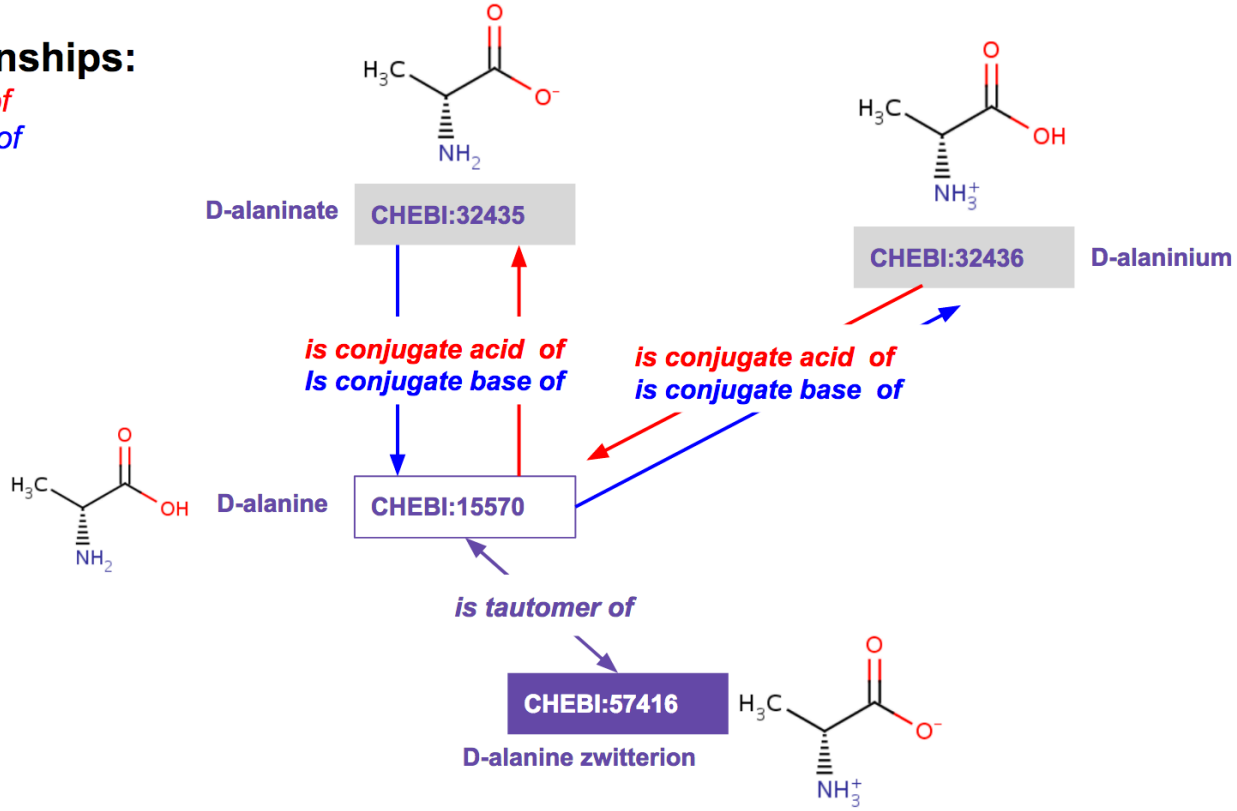
CHEBI:57416

D-alanine zwitterion

Different chemical species for the same compound

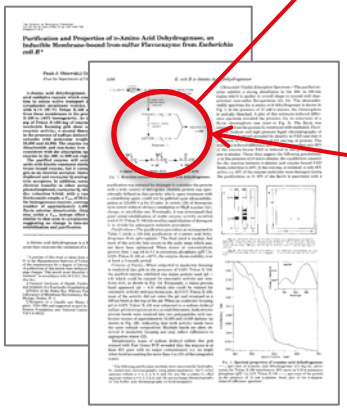
ChEBI relationships:

Is conjugate acid of
Is conjugate base of
Is tautomer of

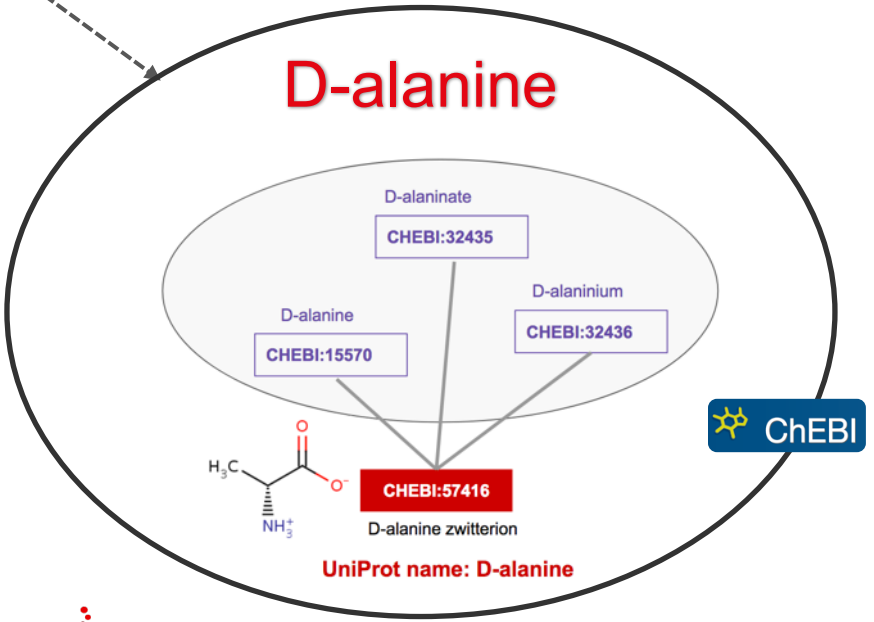


Rhea curation: choice of reaction participants

5,10-methylenetetrahydrofolate + D-alanine + H_2O = tetrahydrofolate + 2-methylserine



Rhea curators

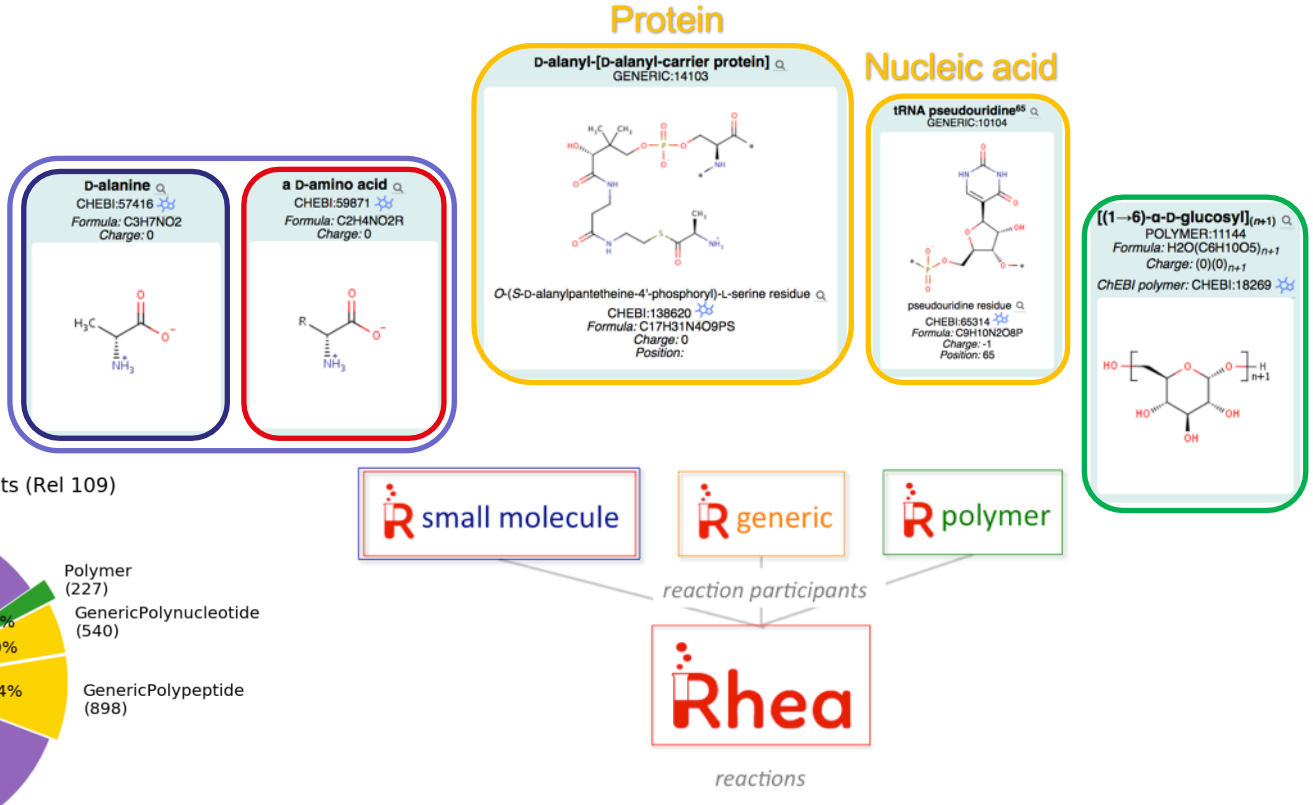


Rhea & MetaCyc: same convention
KEGG: different convention
→ Reconciliation issues (MetaNetX/MNXref)

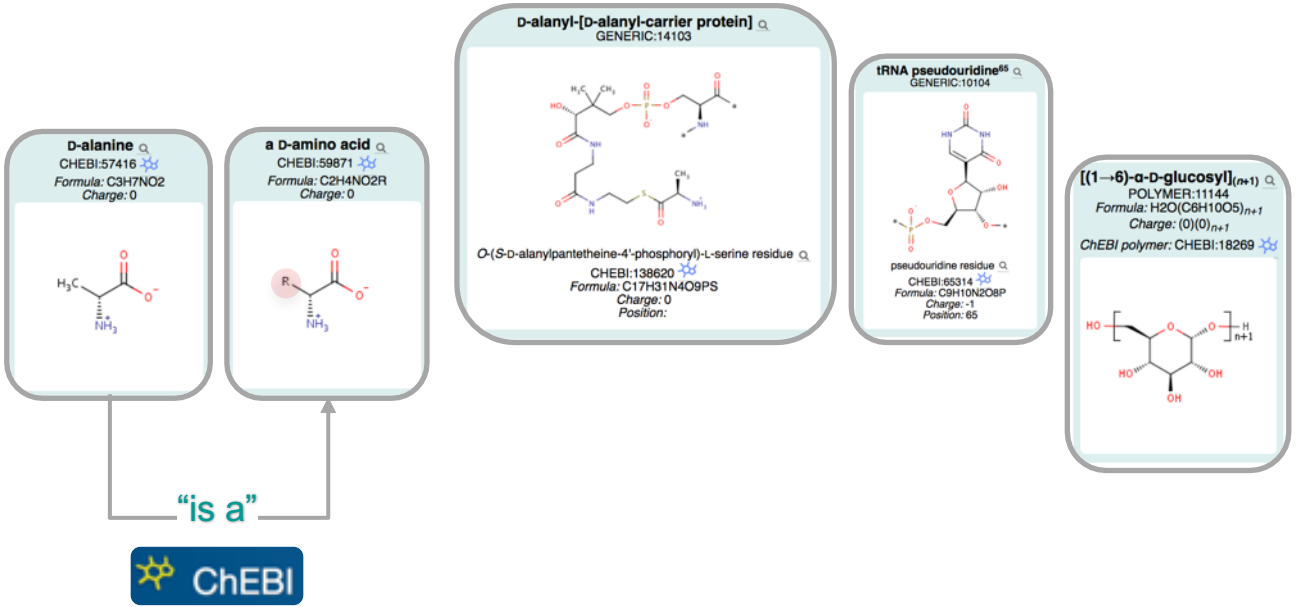
Rhea Major microspecies at pH 7.3

ChemAxon

Rhea reaction participants



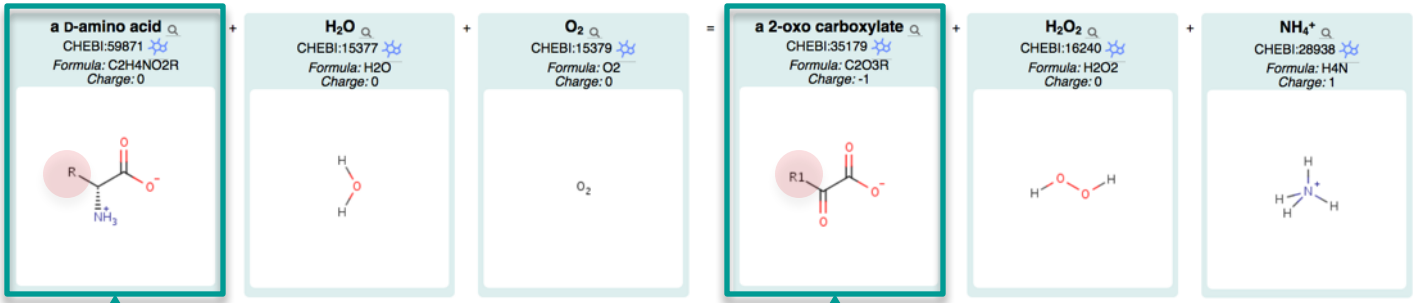
Rhea reactions hierarchical classification (exploit ChEBI ontology)



RHEA:21816 (APPROVED)

a D-amino acid + H2O + O2 = a 2-oxo carboxylate + H2O2 + NH4(+)

Last modified: 2017-07-11. **Chemically balanced:** yes.



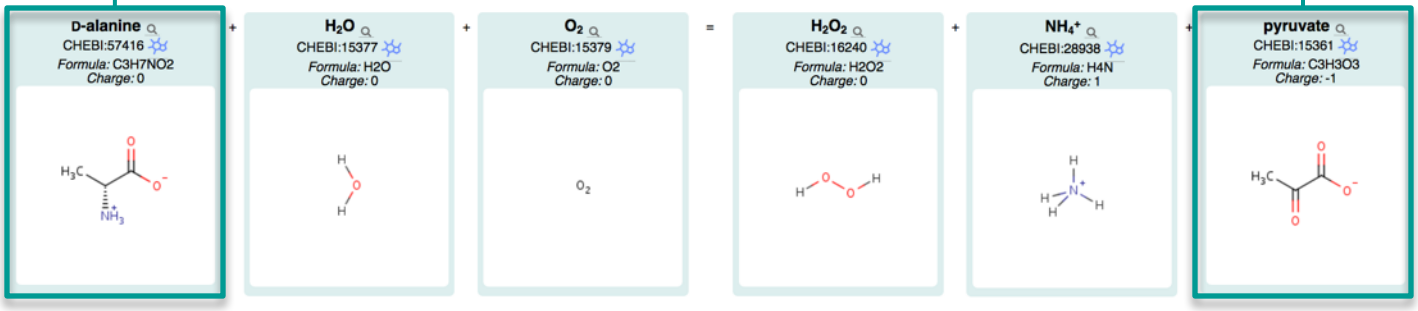
"is a" ChEBI

"is a" ChEBI

RHEA:22688 (APPROVED)

D-alanine + H2O + O2 = H2O2 + NH4(+) + pyruvate

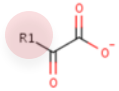
Last modified: 2017-11-08. **Chemically balanced:** yes.



RHEA:21816 (APPROVED)

a D-amino acid + H2O + O2 = a 2-oxo carboxylate + H2O2 + NH4(+)

Last modified: 2017-07-11. **Chemically balanced:** yes.

a D-amino acid	+	H_2O	+	O_2	=	a 2-oxo carboxylate	+	H_2O_2	+	NH_4^+
CHEBI:59871		CHEBI:15377		CHEBI:15379		CHEBI:35179		CHEBI:16240		CHEBI:28938
Formula: C2H4NO2R *Charge:* 0		*Formula:* H2O *Charge:* 0		*Formula:* O2 *Charge:* 0		*Formula:* C2O3R *Charge:* -1		*Formula:* H2O2 *Charge:* 0		*Formula:* H4N *Charge:* 1
										

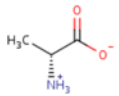
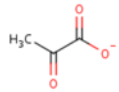
Rhea

"is a" reaction relationship

RHEA:22688 (APPROVED)

D-alanine + H2O + O2 = H2O2 + NH4(+) + pyruvate

Last modified: 2017-11-08. **Chemically balanced:** yes.

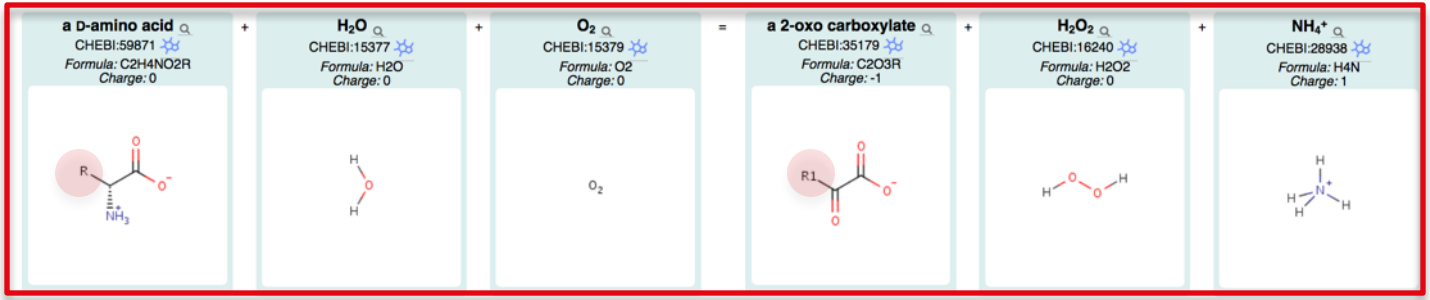
D-alanine	+	H_2O	+	O_2	=	H_2O_2	+	NH_4^+	+	pyruvate
CHEBI:57416		CHEBI:15377		CHEBI:15379		CHEBI:16240		CHEBI:28938		CHEBI:15361
Formula: C3H7NO2 *Charge:* 0		*Formula:* H2O *Charge:* 0		*Formula:* O2 *Charge:* 0		*Formula:* H2O2 *Charge:* 0		*Formula:* H4N *Charge:* 1		*Formula:* C3H3O3 *Charge:* -1
										

< 20% Rhea reactions are classified
(expert curation, on-going process)

RHEA:21816 (APPROVED)

a D-amino acid + H2O + O2 = a 2-oxo carboxylate + H2O2 + NH4(+)

Last modified: 2017-07-11. **Chemically balanced:** yes.



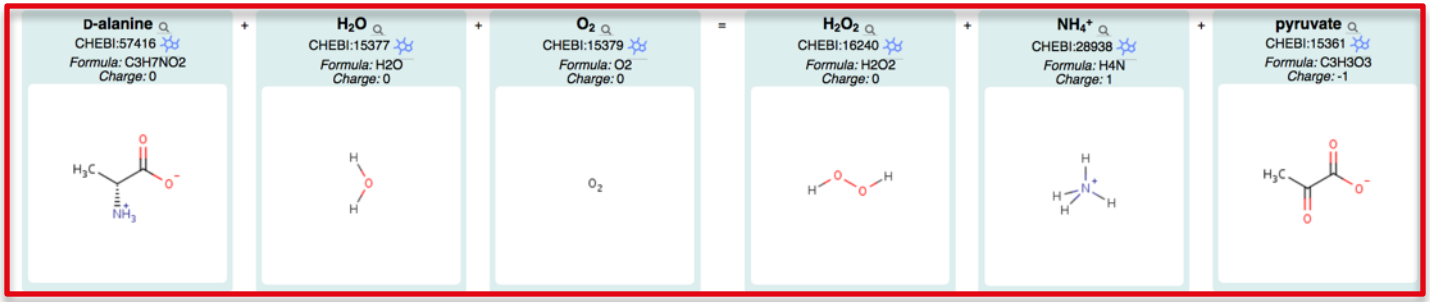
EC 1.4.3.**3** (*D-amino-acid oxidase*)

Rhea
"is a" reaction relationship

RHEA:22688 (APPROVED)

D-alanine + H2O + O2 = H2O2 + NH4(+) + pyruvate

Last modified: 2017-11-08. **Chemically balanced:** yes.



EC 1.4.3.**19** (alternate reaction of *glycine oxidase*)

Rhea reactions and enzyme catalysts

- Enzyme nomenclature (EC numbers)

- UniProt enzymes

Enzyme classification and EC numbers

- **EC 1 Oxidoreductases**
  - EC 1.1 Acting on the CH-OH group of donors
    - EC 1.1.1 With NAD+ or NADP+ as acceptor
      - EC 1.1.1.2 alcohol dehydrogenase (NADP+)

        an alcohol + NADP+ = an aldehyde + NADPH + H+

      - ...
      - EC 1.1.1.19 glucuronate reductase

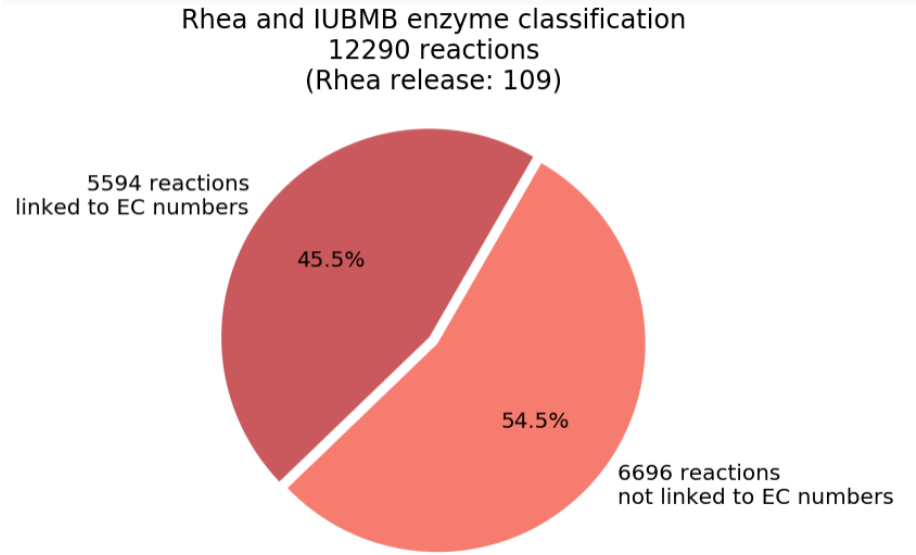
        L-gulonate + NADP+ = D-glucuronate + NADPH + H+

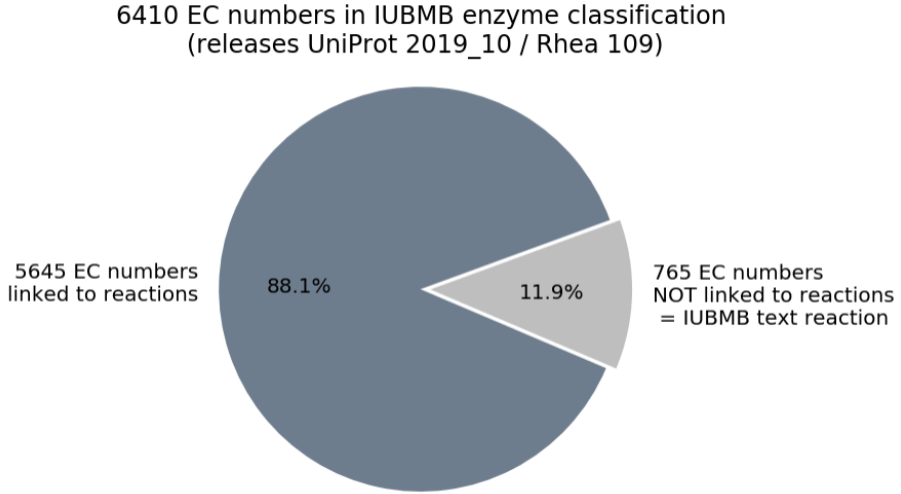
is a

missing deeper levels

- **EC 2 Transferases**
- **EC 3 Hydrolases**
- **EC 4 Lyases**
- **EC 5 Isomerases**
- **EC 6 Ligases**
- **EC 7 Translocases**

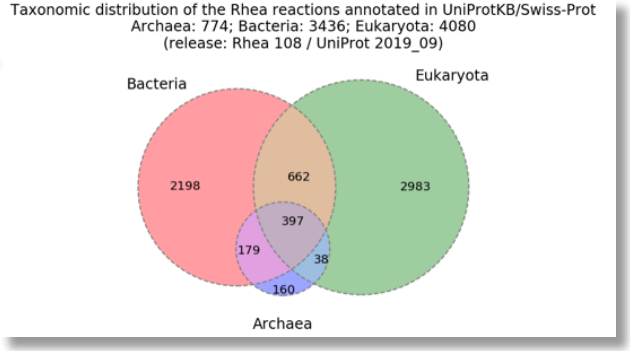
Rhea and Enzyme classification





Rhea is used in UniProt since December 2018

Rhea reactions used in UniProtKB/Swiss-Prot	6’629 / 12’290 (54.1%)
UniProtKB/Swiss-Prot proteins annotated with Rhea reactions	217’207 (38.7%)
UniProtKB/TrEMBL proteins automatically annotated with Rhea reactions	17’141’552 (9.4%)



OXFORD ACADEMIC

Bioinformatics

ACCEPTED MANUSCRIPT

Enzyme annotation in UniProtKB using Rhea

Anne Morgat ✉, Thierry Lombardot, Elisabeth Coudert, Kristian Axelsen, Teresa Batista Neto, Sebastien Gehant, Parit Bansal, Jerven Bolleman, Elisabeth Gasteiger, Edouard de Castro ... Show more

Bioinformatics, btz817, https://doi.org/10.1093/bioinformatics/btz817

Published: 05 November 2019 **Article history** ▾

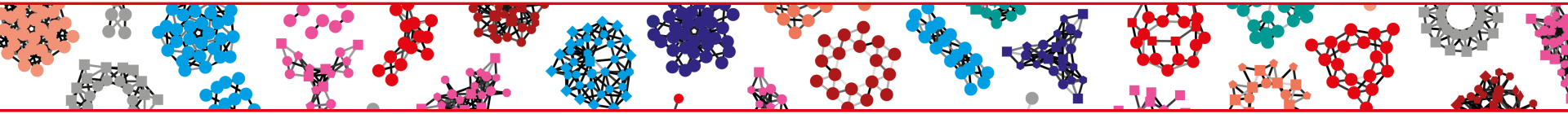
PDF Split View Cite Permissions Share ▾

Abstract

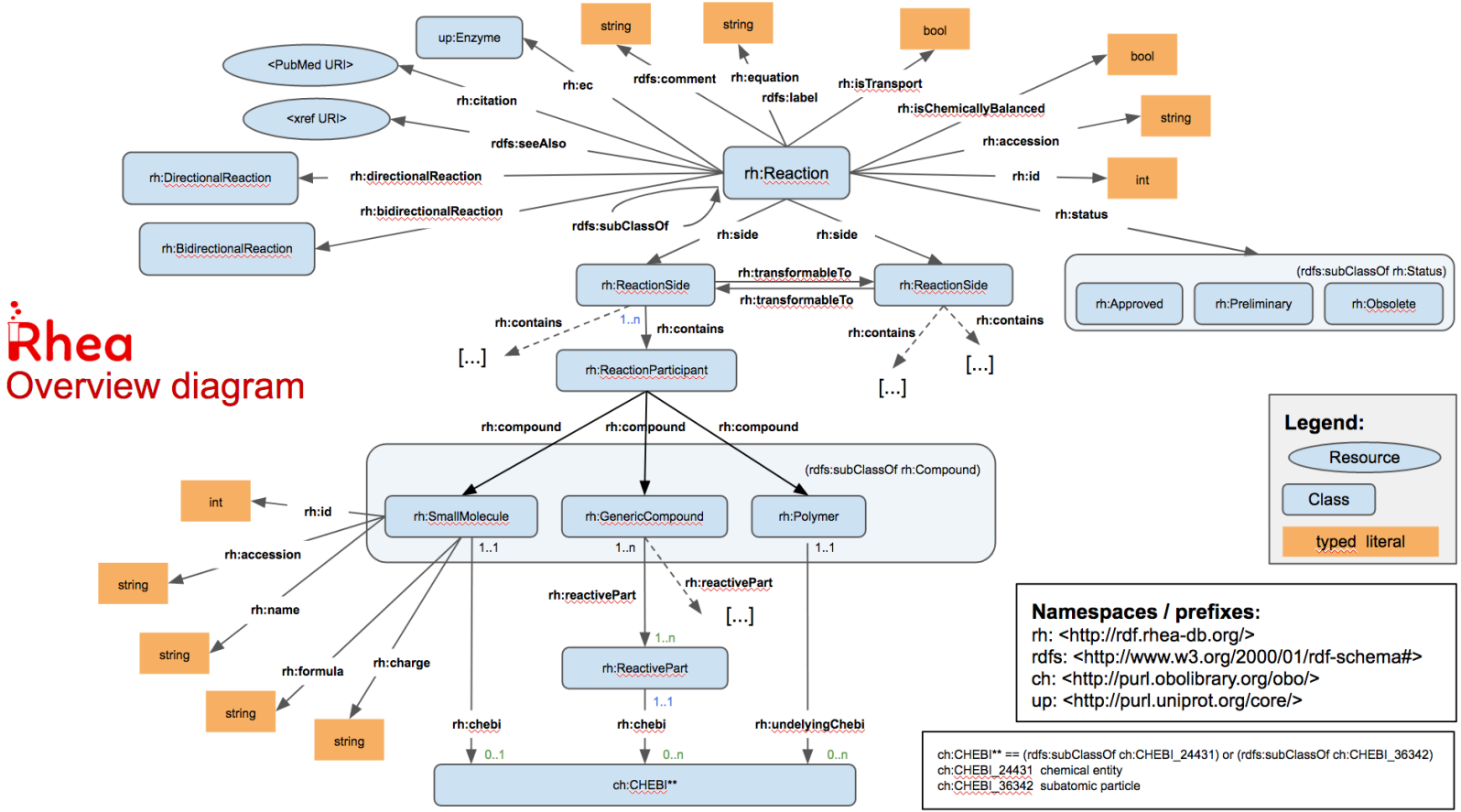
Motivation

To provide high quality computationally tractable enzyme annotation in UniProtKB using Rhea, a comprehensive expert-curated knowledgebase of biochemical reactions which describes reaction participants using the ChEBI (Chemical Entities of Biological Interest) ontology.

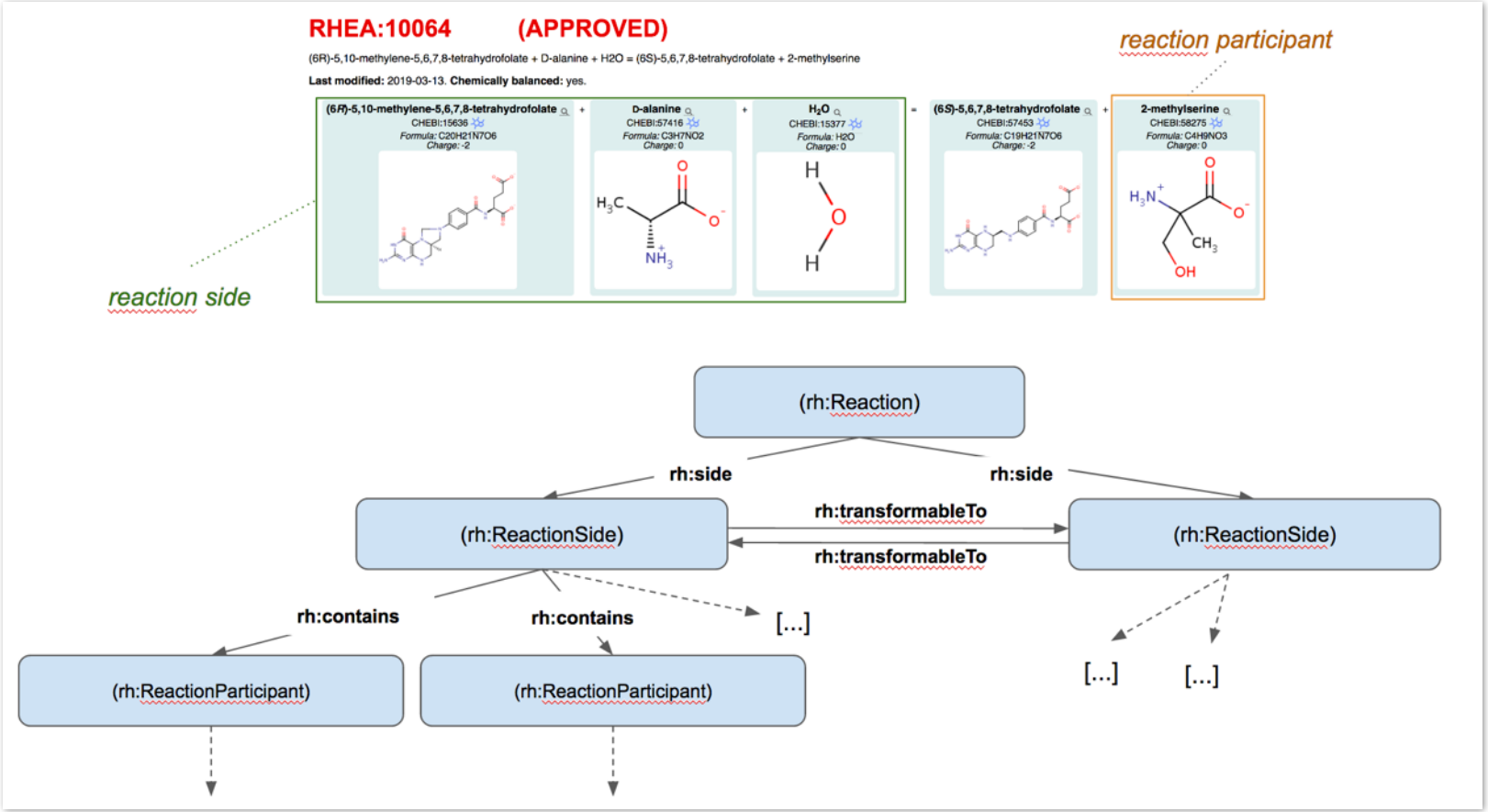
Overview



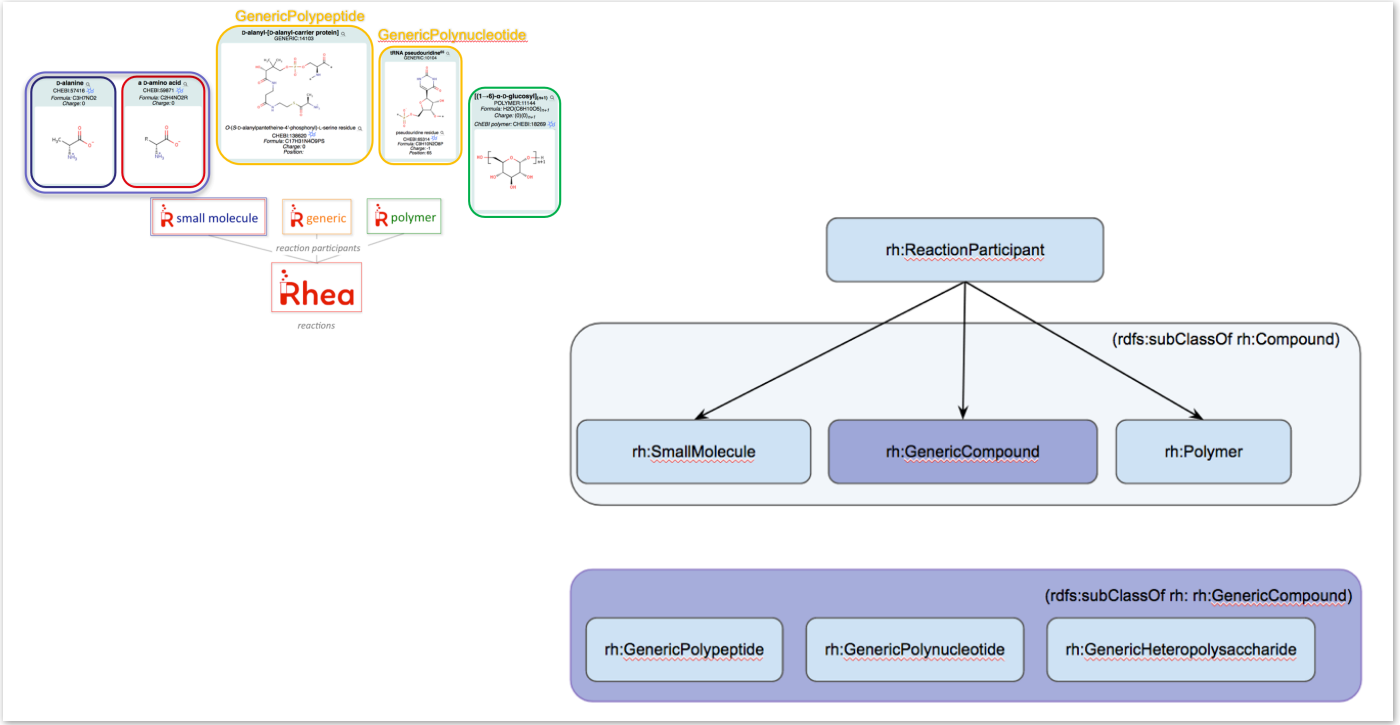
Data model (https://www.rhea-db.org/**rhea_rdf_documentation.pdf**)



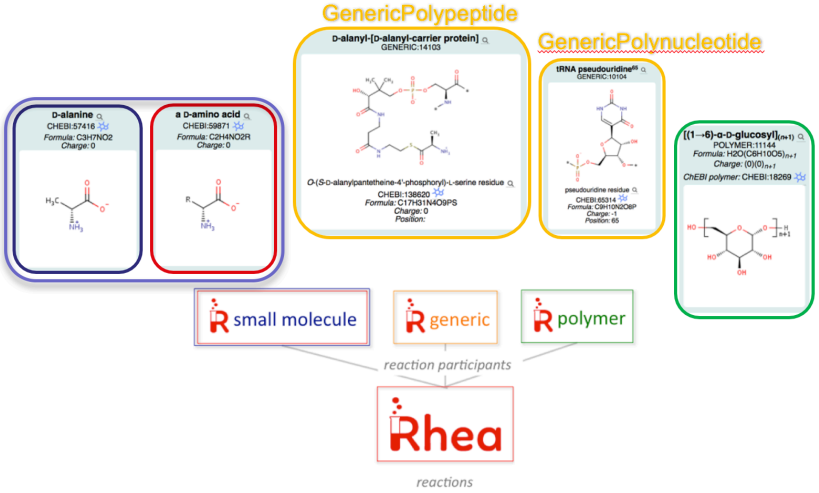
Reactions, sides and participants

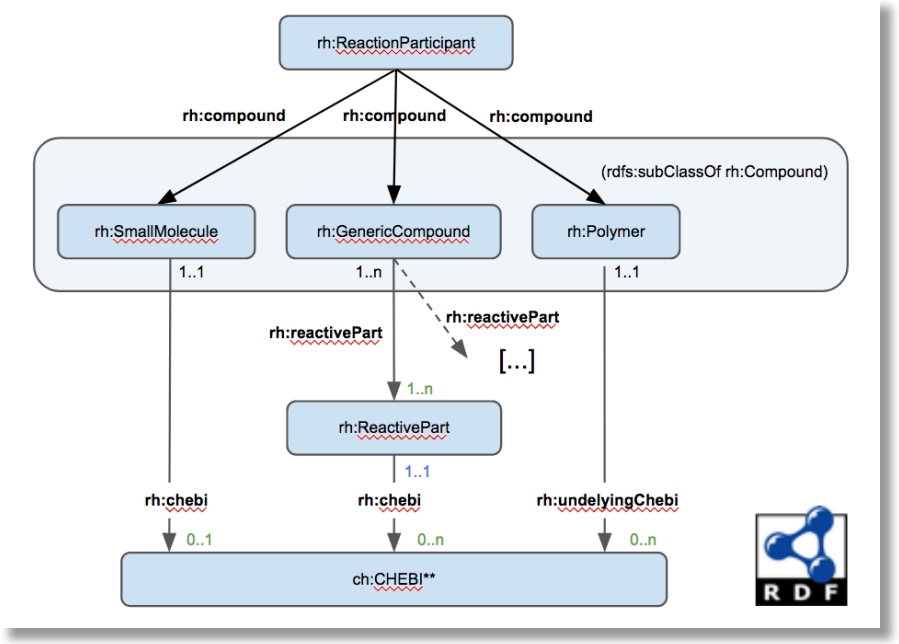


Reactions participants: rh:Compound classes

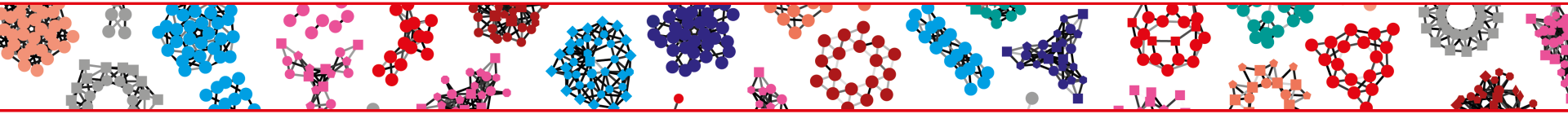


Reactions participants: links to ChEBI





Overview



01 What is Rhea?

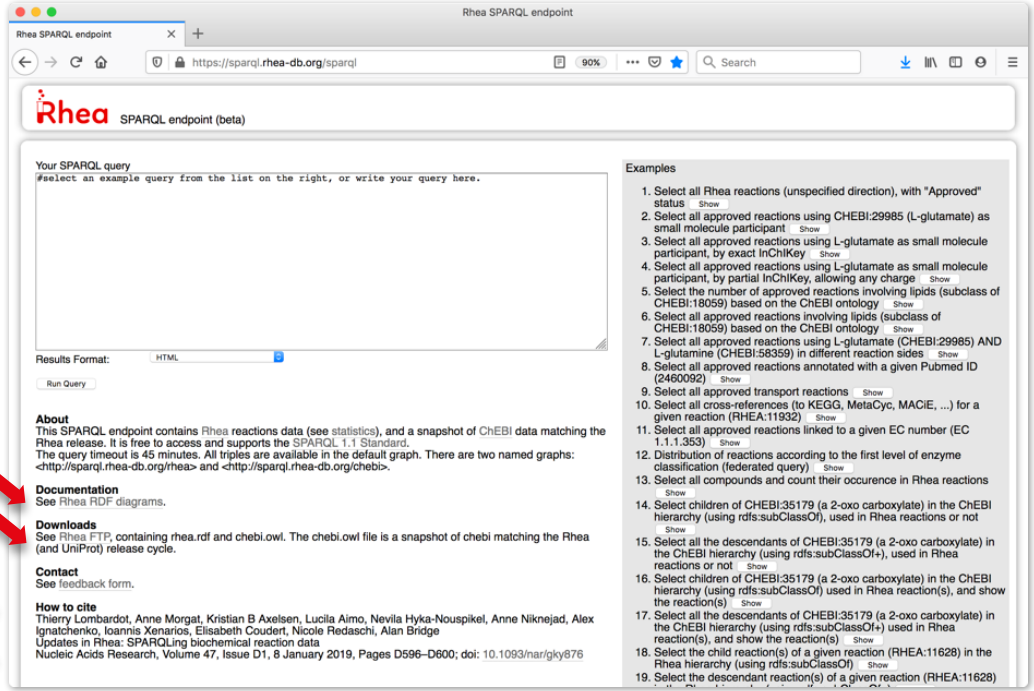
02 Data model

03 SPARQLing Rhea

04 Summary

Rhea SPARQL endpoint: rhea.rdf + chebi.owl

https://sparql.rhea-db.org/sparql



D596–D600 *Nucleic Acids Research, 2019, Vol. 47, Database issue* *Published online 1 October 2018*
doi: 10.1093/nar/gky876

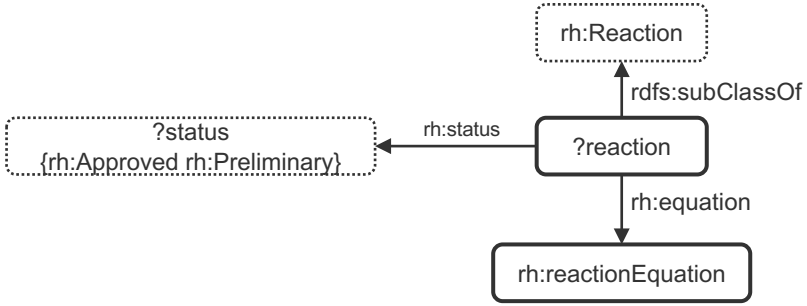
Updates in Rhea: SPARQLing biochemical reaction data

Thierry Lombardot[1,†], Anne Morgat[1,†], Kristian B. Axelsen[1], Lucila Aimo[1], Nevila Hyka-Nouspikel[1], Anne Niknejad[2], Alex Ignatchenko[3], Ioannis Xenarios[4,5], Elisabeth Coudert[1], Nicole Redaschi[1] and Alan Bridge[1,*]

[1]Swiss-Prot Group, SIB Swiss Institute of Bioinformatics, CMU, 1 rue Michel-Servet, CH-1211 Geneva 4, Switzerland, [2]Vital-IT Group, SIB Swiss Institute of Bioinformatics, Quartier Sorge, Bâtiment Génopode, CH-1015 Lausanne, Switzerland, [3]European Molecular Biology Laboratory, European Bioinformatics Institute (EMBL-EBI), Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SD, UK, [4]Department of Biochemistry, University of Geneva, CH-1211 Geneva, Switzerland and [5]Center for Integrative Genomics, University of Lausanne, CH-1015 Lausanne, Switzerland

Received August 13, 2018; Editorial Decision September 11, 2018; Accepted September 27, 2018

Q1: Retrieve **all** Rhea reactions (approved or preliminary) and display their chemical equation




```
#endpoint: https://sparql.rhea-db.org/sparql

PREFIX rh:<http://rdf.rhea-db.org/>
SELECT ?reaction ?reactionEquation
WHERE {
  ?reaction rdfs:subClassOf rh:Reaction .
  ?reaction rh:status ?status .
  VALUES ?status {rh:Approved rh:Preliminary}
  ?reaction rh:equation ?reactionEquation .
}
ORDER BY ?reaction
```

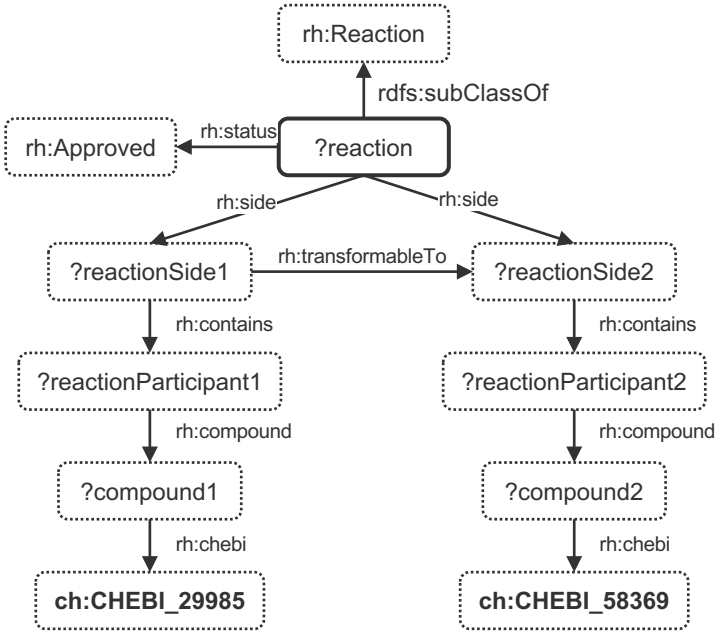
reaction	reactionEquation
http://rdf.rhea-db.org/10000	"H2O + pentanamide = NH4(+) + pentanoate"
http://rdf.rhea-db.org/10004	"benzyl isothiocyanate = benzyl thiocyanate"
http://rdf.rhea-db.org/10008	"[protein]-dithiol + a hydroperoxide = [protein]-disulfide + an alcohol + H2O"
http://rdf.rhea-db.org/10012	"(R)-6-hydroxynicotine + H2O + O2 = 6-hydroxypseudooxynicotine + H2O2"
http://rdf.rhea-db.org/10016	"H2O + O-sinapoylcholine = choline + H(+) + trans-sinapate"
http://rdf.rhea-db.org/10020	"H2O + L-saccharopine + NADP(+) = (S)-2-amino-6-oxohexanoate + H(+) + L-glutamate + NADPH"
http://rdf.rhea-db.org/10024	"L-lysyl-[histone] + S-adenosyl-L-methionine = H(+) + N(6)-methyl-L-lysyl-[histone] + S-adenosyl-L-homocysteine"
http://rdf.rhea-db.org/10028	"D-glutamate + H2O + O2 = 2-oxoglutarate + H2O2 + NH4(+)"
http://rdf.rhea-db.org/10032	"[protein]-C-terminal-L-glutamine + H2O = [protein]-C-terminal-L-glutamate + NH4(+)"
http://rdf.rhea-db.org/10036	"4-hydroxy-3-methoxy-benzenemethanol + O2 = 4-hydroxy-3-methoxybenzaldehyde + H2O2"

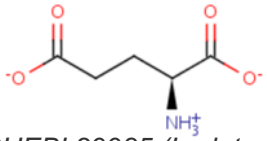
[…]

Chemical search (data from chebi.owl)

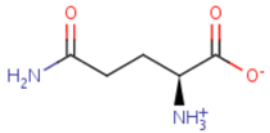
- Identifiers (**ChEBI IDs**, cross-references IDs) *[query Q2]*
- Names and synonyms
- **ChEBI ontology ('is a', 'has role' relationships)** *[query Q3]*
- **Structural search**
  - Formula, SMILES, InChi, **InChiKey** *[query Q4]*
  - ***Substructure search (IDSM/Sachem)*** *[query Q20]*

Q2: Retrieve all approved reactions using L-glutamate (CHEBI:29985) AND L-glutamine (CHEBI:58359) in opposite reaction sides





CHEBI:29985 (L-glutamate)

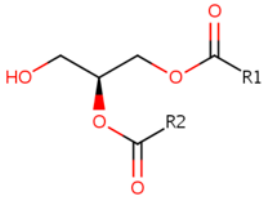


CHEBI:58359 (L-glutamine)

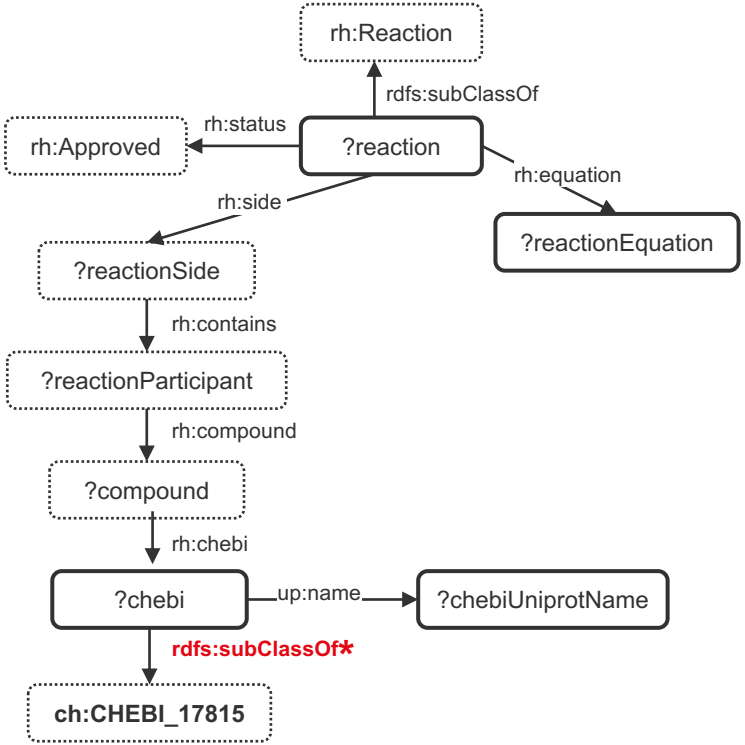
```
#endpoint:https://sparql.rhea-db.org/sparql

PREFIX rh:<http://rdf.rhea-db.org/>
PREFIX ch:<http://purl.obolibrary.org/obo/>

SELECT ?reaction
WHERE {
  ?reaction rdfs:subClassOf rh:Reaction .
  ?reaction rh:status rh:Approved .

  ?reaction rh:side ?reactionSide1 .
  ?reactionSide1 rh:contains ?participant1 .
  ?participant1 rh:compound ?compound1 .
  ?compound1 rh:chebi ch:CHEBI_29985 .

  ?reaction rh:side ?reactionSide2 .
  ?reactionSide2 rh:contains ?participant2 .
  ?participant2 rh:compound ?compound2 .
  ?compound2 rh:chebi ch:CHEBI_58359 .

  ?reactionSide1 rh:transformableTo ?reactionSide2 .
}
```

reaction
http://rdf.rhea-db.org/21732
http://rdf.rhea-db.org/23256
http://rdf.rhea-db.org/24384
http://rdf.rhea-db.org/24793
http://rdf.rhea-db.org/25512
http://rdf.rhea-db.org/26289
http://rdf.rhea-db.org/26426
http://rdf.rhea-db.org/31507
http://rdf.rhea-db.org/34547
http://rdf.rhea-db.org/34551
http://rdf.rhea-db.org/11672
http://rdf.rhea-db.org/11680
http://rdf.rhea-db.org/12128
http://rdf.rhea-db.org/12228
http://rdf.rhea-db.org/12544
http://rdf.rhea-db.org/13237
http://rdf.rhea-db.org/13753
http://rdf.rhea-db.org/14513
http://rdf.rhea-db.org/14905
http://rdf.rhea-db.org/16169
http://rdf.rhea-db.org/17129
http://rdf.rhea-db.org/17521
http://rdf.rhea-db.org/18633
http://rdf.rhea-db.org/15501
http://rdf.rhea-db.org/15889
http://rdf.rhea-db.org/52896
http://rdf.rhea-db.org/54084
http://rdf.rhea-db.org/57928

Q3: Select all approved reactions with **CHEBI:17815 (a 1,2-diacyl-sn-glycerol)** or one of its descendant.



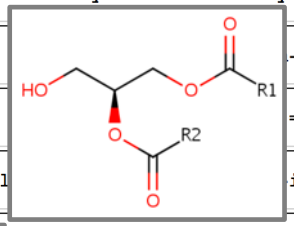
CHEBI:17815
(a 1,2-diacyl-sn-glycerol)




```
#endpoint:https://sparql.rhea-db.org/sparql

PREFIX up:<http://purl.uniprot.org/core/>

SELECT distinct ?chebi ?chebiUniprotName
                ?reaction ?reactionEquation
WHERE {
  ?reaction rdfs:subClassOf rh:Reaction .
  ?reaction rh:status rh:Approved .
  ?reaction rh:equation ?reactionEquation .
  ?reaction rh:side ?reactionSide .
  ?reactionSide rh:contains ?participant .
  ?participant rh:compound ?compound .
  ?compound rh:chebi ?chebi .
  ?chebi rdfs:subClassOf* ch:CHEBI_17815 .
  ?chebi up:name ?chebiUniprotName .
}
```

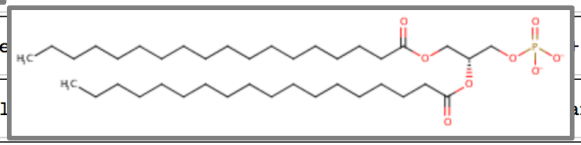
Q3: Select all approved reactions with **CHEBI:17815 (a 1,2-diacyl-sn-glycerol)** or one of its descendant.

reaction	reactionEquation
http://rdf.rhea-db.org/10272	"a 1,2-diacyl-sn-glycerol + ATP = a 1,2-diacyl-sn-glycero-3-phosphate + ADP + H(+)"
http://rdf.rhea-db.org/10604	"a 1,2-diacyl-sn-glycero-3-phosphocholine + H2O = a 1,2-diacyl-sn-glycerol + H(+) + phosphocholine"
http://rdf.rhea-db.org/10868	"a 1,2-diacyl-sn-glycerol + an acyl-CoA = a triacyl-sn-glycerol + CoA"
http://rdf.rhea-db.org/13301	"a 1,2-diacyl-sn-glycerol -glycerol + a sterol ester"
http://rdf.rhea-db.org/14057	"a 1,2-diacyl-sn-glycerol = a monoacylglycerophospholipid + a triacyl-sn-glycerol"
http://rdf.rhea-db.org/14333	"a 6-(alpha-D-glucosaminyl inositol = 6-(alpha-D-glucosaminyl)-1D-myo-inositol 1,2-cyclic phosphate + a 1,2-diacyl-sn-glycerol"
http://rdf.rhea-db.org/14945	"a 1,2-diacyl-sn-glycerol + UDP-alpha-D-galactose = a 1,2-diacyl-3-O-(beta-D-galactosyl)-sn-glycerol + H(+) + UDP"



[...]

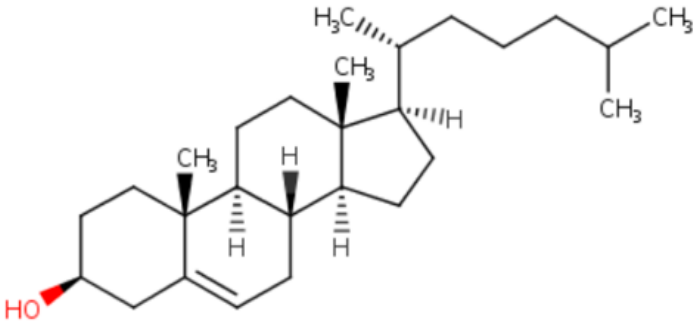
http://rdf.rhea-db.org/33335	"1,2-dioctadecanoyl-sn-glycero-3-phosphate + H2O = 1,2-dioctadecanoyl-sn-glycerol + phosphate"
http://rdf.rhea-db.org/36179	"1-O-alkyl-2-acyl-sn-glycerol + CDP-choline - CMP + H(+)"
http://rdf.rhea-db.org/36187	"1-O-alkyl-2-acyl-sn-glycerol + CDP-ethanol anolamine + CMP + H(+)"
http://rdf.rhea-db.org/36207	"1-O-(1Z-alkenyl)-sn-glycero-3-phosphate + an acyl-CoA = 1-O-(1Z-alkenyl)-2-acyl-sn-glycero-3-phosphate + CoA"



[...]

Chemical structure search

CHEBI:16113
(cholesterol)



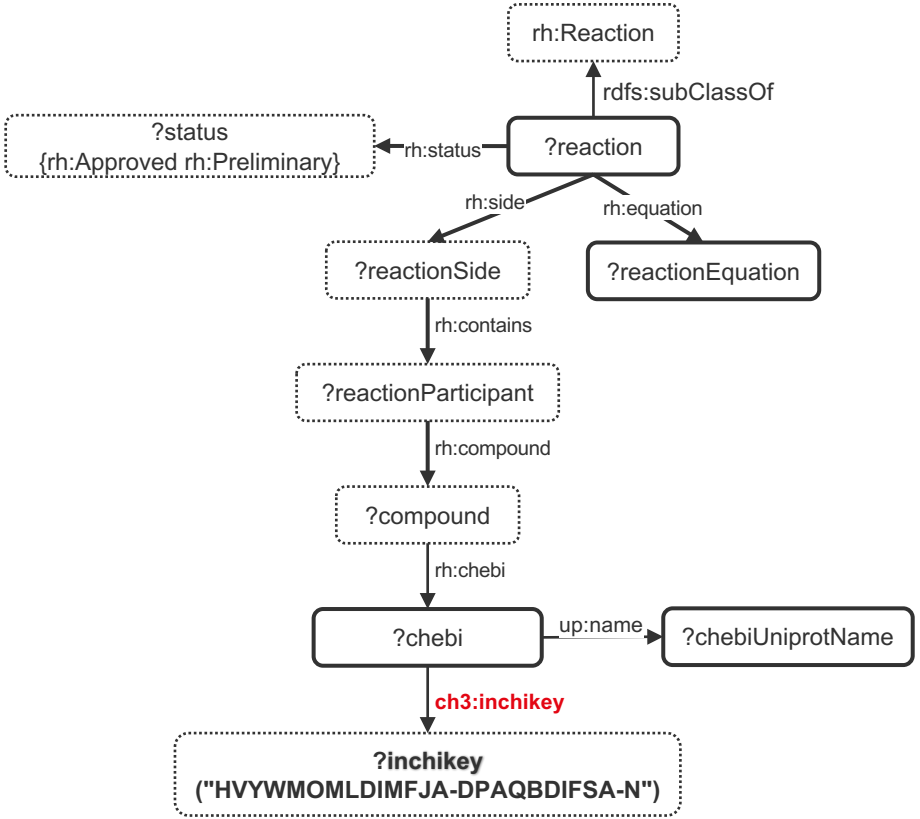
Formula: C27H46O

SMILES
C1[C@@]2([C@]3(CC[C@]4([C@]([C@@]3(CC=C2C[C@H](C1)O)[H])(CC[C@@]4([C@H](C)CCCC(C)C)[H])[H])C)[H])C

InChI
InChI=1S/C27H46O/c1-18(2)7-6-8-19(3)23-11-12-24-22-10-9-20-17-21(28)13-15-26(20,4)25(22)14-16-27(23,24)5/h9,18-19,21-25,28H,6-8,10-17H2,1-5H3/t19-,21+,22+,23-,24+,25+,26+,27-/m1/s1

InChIKey
HVYWMOMLDIMFJA-DPAQBDIFSA-N

Q4: Retrieve Rhea reactions that involve cholesterol using its **InChiKey**




```
#endpoint:https://sparql.rhea-db.org/sparql

PREFIX rh:<http://rdf.rhea-db.org/>
PREFIX ch:<http://purl.obolibrary.org/obo/>
PREFIX ch3:<http://purl.obolibrary.org/obo/chebi/>
PREFIX up:<http://purl.uniprot.org/core/>

SELECT distinct ?chebi ?chebiUniprotName
            ?reaction ?reactionEquation
WHERE {
  ?reaction rdfs:subClassOf rh:Reaction .
  ?reaction rh:status ?status .
  VALUES (?status) {(rh:Approved) (rh:Preliminary)}
  ?reaction rh:equation ?reactionEquation .
  ?reaction rh:side ?reactionSide .
  ?reactionSide rh:contains ?participant .
  ?participant rh:compound ?compound .
  ?compound rh:chebi ?chebi .
  ?chebi up:name ?chebiUniprotName .
  ?chebi ch3:inchikey ?inchikey .
  VALUES (?inchikey) {("HVYWMOMLDIMFJA-DPAQBDIFSA-N")}
}
ORDER BY ?reaction
```

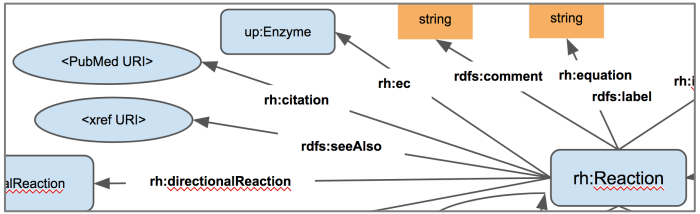
Q4: Retrieve Rhea reactions that involve cholesterol using its **InChiKey**

chebi	chebiUniprotName	reaction	reactionEquation
http://purl.obolibrary.org/obo/CHEBI_16113	"cholesterol"	http://rdf.rhea-db.org/11956	"cholesteryl-beta-D-glucoside + H2O = cholesterol + D-glucose"
http://purl.obolibrary.org/obo/CHEBI_16113	"cholesterol"	http://rdf.rhea-db.org/17729	"an acyl-CoA + cholesterol = a cholesterol ester + CoA"
http://purl.obolibrary.org/obo/CHEBI_16113	"cholesterol"	http://rdf.rhea-db.org/21104	"AH2 + cholesterol + O2 = 25-hydroxycholesterol + A + H2O"
http://purl.obolibrary.org/obo/CHEBI_16113	"cholesterol"	http://rdf.rhea-db.org/21328	"cholesterol + O2 = cholest-4-en-3-one + H2O2"
http://purl.obolibrary.org/obo/CHEBI_16113	"cholesterol"	http://rdf.rhea-db.org/21812	"cholesterol + O2 + reduced [NADPH--hemoprotein reductase] = 7alpha-hydroxycholesterol + H(+) + H2O + oxidized [NADPH--hemoprotein reductase]"
http://purl.obolibrary.org/obo/CHEBI_16113	"cholesterol"	http://rdf.rhea-db.org/22716	"cholesterol + O2 + reduced [NADPH--hemoprotein reductase] = (24S)-hydroxycholesterol + H(+) + H2O + oxidized [NADPH--hemoprotein reductase]"
http://purl.obolibrary.org/obo/CHEBI_16113	"cholesterol"	http://rdf.rhea-db.org/23984	"cholesterol + NADP(+) = cholesta-5,7-dien-3beta-ol + H(+) + NADPH"
http://purl.obolibrary.org/obo/CHEBI_16113	"cholesterol"	http://rdf.rhea-db.org/32183	"cholesterol + O2 = cholest-5-en-3-one + H2O2"
http://purl.obolibrary.org/obo/CHEBI_16113	"cholesterol"	http://rdf.rhea-db.org/33875	"cholesteryl (9Z-octadecenoate) + H2O = (9Z)-octadecenoate + cholesterol + H(+)"
http://purl.obolibrary.org/obo/CHEBI_16113	"cholesterol"	http://rdf.rhea-db.org/34087	"a 1,2-diacyl-sn-glycero-3-phosphocholine + cholesterol = a 1-acyl-sn-glycero-3-phosphocholine + a cholesterol ester"

[…]

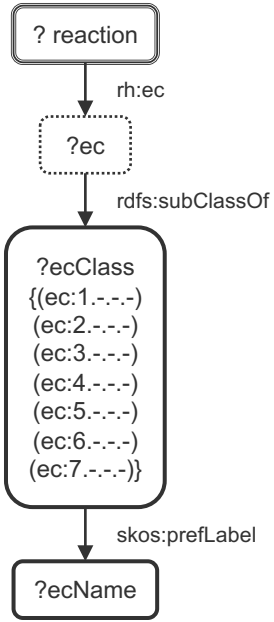
Enzyme search

- Enzyme nomenclature (EC numbers)



- UniProt enzymes

Q5: Distribution of Rhea reactions based on the first level of IUBMB enzyme classification



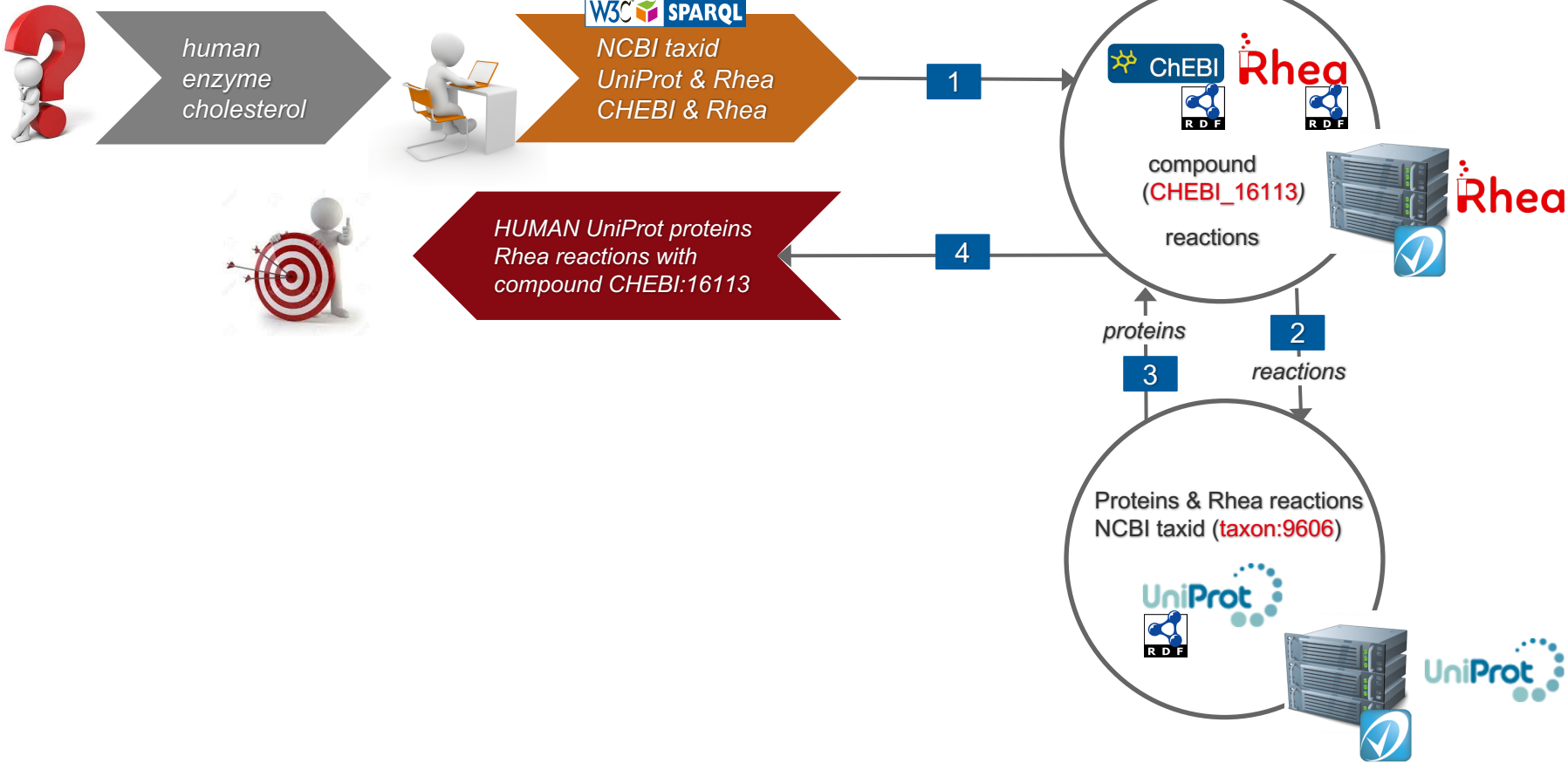

```
#endpoint:https://sparql.rhea-db.org/sparql

PREFIX rh:<http://rdf.rhea-db.org/>
PREFIX ec:<http://purl.uniprot.org/enzyme/>
PREFIX rdfs:<http://www.w3.org/2000/01/rdf-schema#>
PREFIX skos:<http://www.w3.org/2004/02/skos/core#>

SELECT ?ecClass
       ?ecClassName
       (count(?reaction) as ?reactionCount)
WHERE
{
 SERVICE <http://sparql.uniprot.org/sparql> {
   ?ec rdfs:subClassOf ?ecClass .
   ?ecClass skos:prefLabel ?ecClassName .
   VALUES (?ecClass) { (ec:1.-.-.-) (ec:2.-.-.-)(ec:3.-.-.-)
                       (ec:4.-.-.-)(ec:5.-.-.-) (ec:6.-.-.-)
                       (ec:7.-.-.-) }
 }
 ?reaction rh:ec ?ec .
}
ORDER BY ?ecClass
```

ecClass	ecClassName	reactionCount
http://purl.uniprot.org/enzyme/1.-.-.-	Oxidoreductases	2445
http://purl.uniprot.org/enzyme/2.-.-.-	Transferases	2218
http://purl.uniprot.org/enzyme/3.-.-.-	Hydrolases	909
http://purl.uniprot.org/enzyme/4.-.-.-	Lyases	760
http://purl.uniprot.org/enzyme/5.-.-.-	Isomerases	322
http://purl.uniprot.org/enzyme/6.-.-.-	Ligases	224
http://purl.uniprot.org/enzyme/7.-.-.-	Translocases	76

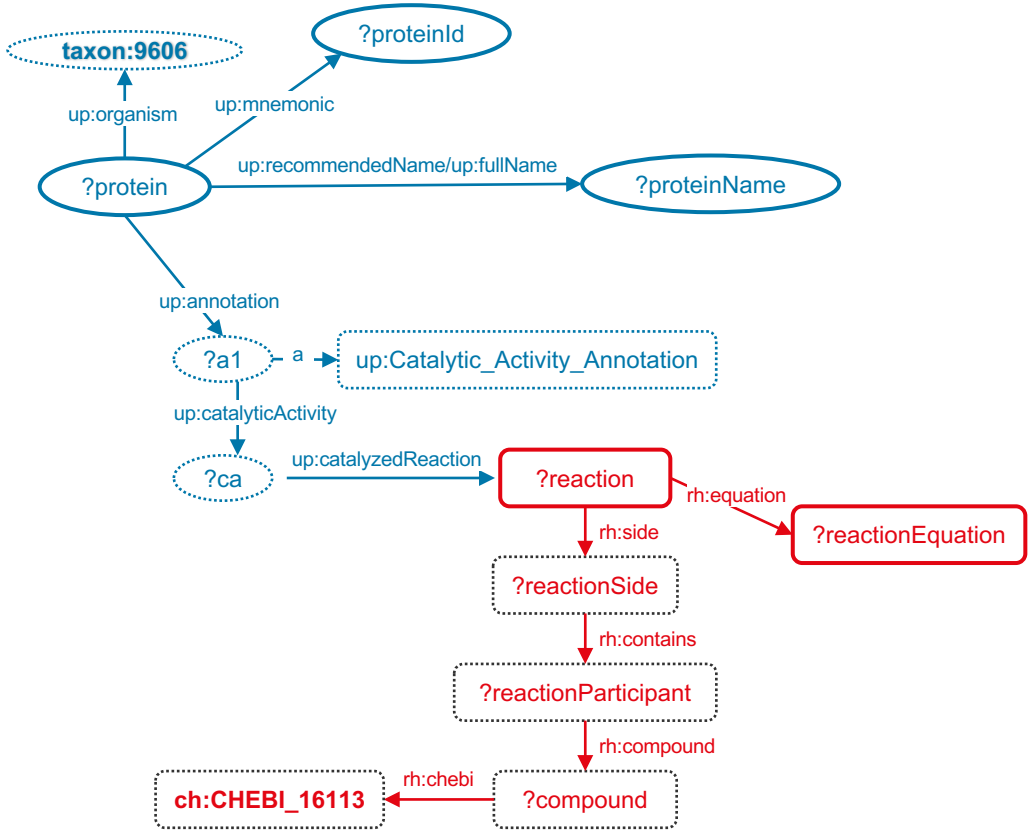
Query: retrieve human enzyme metabolizing cholesterol and their reactions

human
enzyme
cholesterol

Query: retrieve human enzyme metabolizing cholesterol and their reactions



Q6: Retrieve human enzyme metabolizing cholesterol and their reactions




```
#endpoint:https://sparql.rhea-db.org/sparql
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX up: <http://purl.uniprot.org/core/>
PREFIX taxon:<http://purl.uniprot.org/taxonomy/>
PREFIX rh: <http://rdf.rhea-db.org/>
PREFIX ch:<http://purl.obolibrary.org/obo/>
SELECT DISTINCT ?protein  ?proteinId ?proteinName
                ?reaction ?reactionEquation
WHERE {
 ?reaction rdfs:subClassOf rh:Reaction .
 ?reaction rh:side ?reactionSide .
 ?reactionSide rh:contains ?participant .
 ?participant rh:compound ?compound .
 # cholesterol (CHEBI:16113)
 ?compound rh:chebi ch:CHEBI_16113 .
 ?reaction rh:equation ?reactionEquation .
 SERVICE <https://sparql.uniprot.org/sparql/> {
  ?protein up:mnemonic ?proteinId .
  ?protein up:recommendedName ?rn .
  ?rn up:fullName ?proteinName .
  # HUMAN proteins (taxid=9606)
  ?protein up:organism taxon:9606 .
  ?protein up:annotation ?a1 .
  ?a1 a up:Catalytic_Activity_Annotation .
  ?a1 up:catalyticActivity ?ca .
  ?ca up:catalyzedReaction ?reaction .
 }
}
```

Q6: Retrieve human enzyme metabolizing cholesterol and their reactions

protein	proteinId	proteinName	reaction	reactionEquation
http://purl.uniprot.org/uniprot/Q9UBM7	"DHCR7_HUMAN"	"7-dehydrocholesterol reductase"	http://rdf.rhea-db.org/23984	"cholesterol + NADP(+) = cholesta-5,7-dien-3beta-ol + H(+) + NADPH"
http://purl.uniprot.org/uniprot/P45844	"ABCG1_HUMAN"	"ATP-binding cassette sub-family G member 1"	http://rdf.rhea-db.org/39051	"ATP + cholesterol(in) + H2O = ADP + cholesterol(out) + H(+) + phosphate"
http://purl.uniprot.org/uniprot/Q9H222	"ABCG5_HUMAN"	"ATP-binding cassette sub-family G member 5"	http://rdf.rhea-db.org/39051	"ATP + cholesterol(in) + H2O = ADP + cholesterol(out) + H(+) + phosphate"
http://purl.uniprot.org/uniprot/Q9H221	"ABCG8_HUMAN"	"ATP-binding cassette sub-family G member 8"	http://rdf.rhea-db.org/39051	"ATP + cholesterol(in) + H2O = ADP + cholesterol(out) + H(+) + phosphate"
http://purl.uniprot.org/uniprot/P19835	"CEL_HUMAN"	"Bile salt-activated lipase"	http://rdf.rhea-db.org/33875	"cholesteryl (9Z-octadecenoate) + H2O = (9Z)-octadecenoate + cholesterol + H(+)"
http://purl.uniprot.org/uniprot/Q9Y6A2	"CP46A_HUMAN"	"Cholesterol 24-hydroxylase"	http://rdf.rhea-db.org/22716	"cholesterol + O2 + reduced [NADPH--hemoprotein reductase] = (24S)-hydroxycholesterol + H(+) +
http://purl.uniprot.org/uniprot/O95992	"CH25H_HUMAN"	"Cholesterol 25-hydroxylase"	http://rdf.rhea-db.org/21104	"AH2 + cholesterol + O2 = 25-hydroxycholesterol + A + H2O"
http://purl.uniprot.org/uniprot/O95992	"CH25H_HUMAN"	"Cholesterol 25-hydroxylase"	http://rdf.rhea-db.org/46132	"cholesterol + H(+) + NADPH + O2 = 25-hydroxycholesterol + H2O + NADP(+)"
http://purl.uniprot.org/uniprot/P05108	"CP11A_HUMAN"	"Cholesterol side-chain cleavage enzyme, mitochondrial"	http://rdf.rhea-db.org/34335	"cholesterol + 2 H(+) + O2 + 2 reduced [adrenodoxin] = 22R-hydroxycholesterol + H2O + 2 oxidiz
http://purl.uniprot.org/uniprot/P05108	"CP11A_HUMAN"	"Cholesterol side-chain cleavage enzyme, mitochondrial"	http://rdf.rhea-db.org/35739	"cholesterol + 6 H(+) + 3 O2 + 6 reduced [adrenodoxin] = 4-methylpentanal + 4 H2O + 6 oxidized

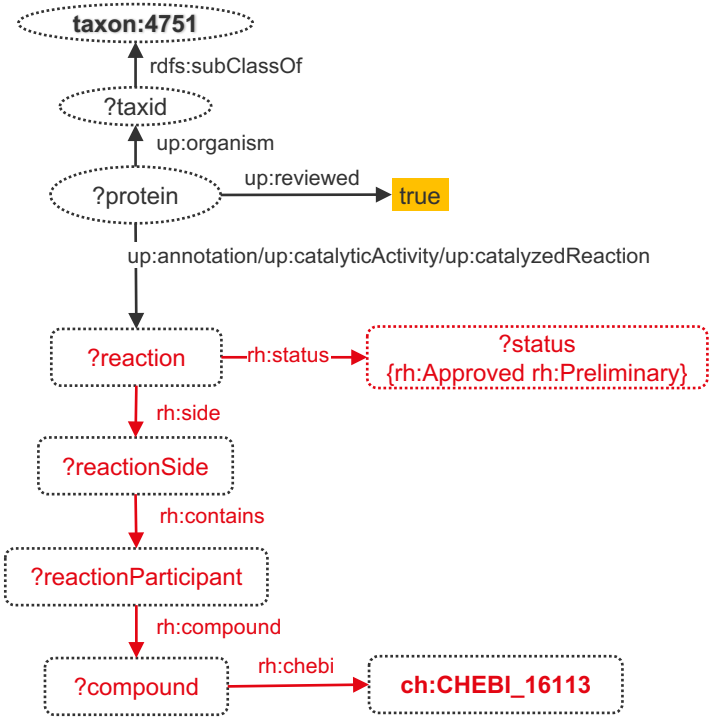
[...]

SPARQL query forms

Query form	
SELECT	Returns all, or a subset of, the variables bound in a query pattern match.
ASK	Returns a boolean indicating whether a query pattern matches or not.
CONSTRUCT	Returns an RDF graph constructed by substituting variables in a set of triple templates.
DESCRIBE	Returns an RDF graph that describes the resources found.

(from https://www.w3.org/TR/rdf-sparql-query/#QueryForms)

Q7: Ask if there are *fungal* UniProtKB/Swiss-Prot protein(s) metabolizing cholesterol (for QC)

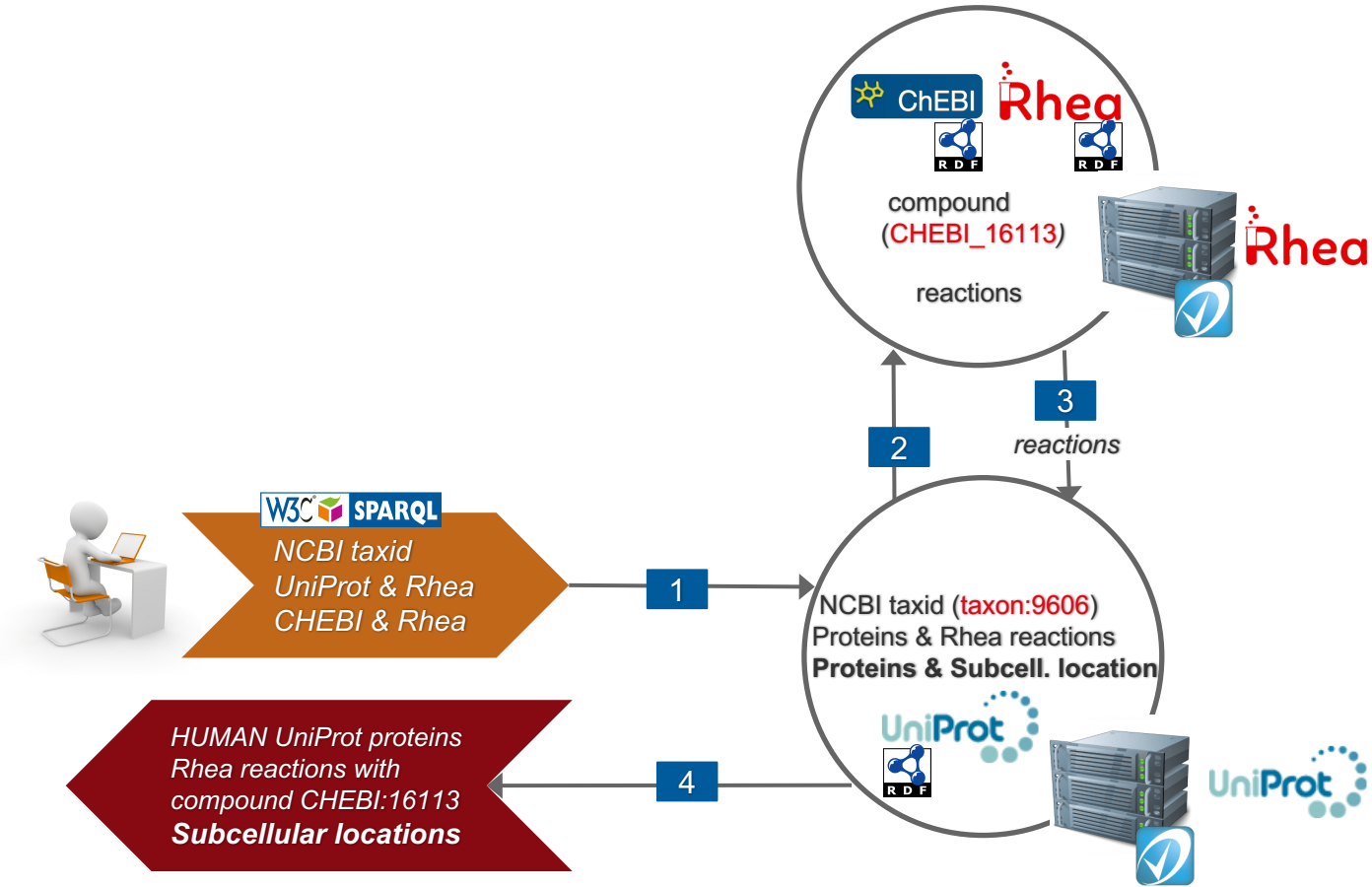



```
#endpoint: https://sparql.uniprot.org/sparql/

PREFIX up:<http://purl.uniprot.org/core/>
PREFIX rh:<http://rdf.rhea-db.org/>
PREFIX taxon:<http://purl.uniprot.org/taxonomy/>
PREFIX ch:<http://purl.obolibrary.org/obo/>
ASK {
 SERVICE <https://sparql.rhea-db.org/sparql> {
  ?reaction rh:status ?status .
  VALUES (?status) {(rh:Approved) (rh:Preliminary)}
  ?reaction rh:side ?reactionSide .
  ?reactionSide rh:contains ?participant .
  ?participant rh:compound ?compound .
  # cholesterol: CHEBI:16113
  ?compound rh:chebi ch:CHEBI_16113 .
 }
 ?protein up:annotation/up:catalyticActivity/up:catalyzedReaction ?reaction .
 # Fungi (taxon:4751),  kingdom
 ?protein up:organism ?taxid .
 ?taxid rdfs:subClassOf taxon:4751 .
 ?protein up:reviewed true .
}
```

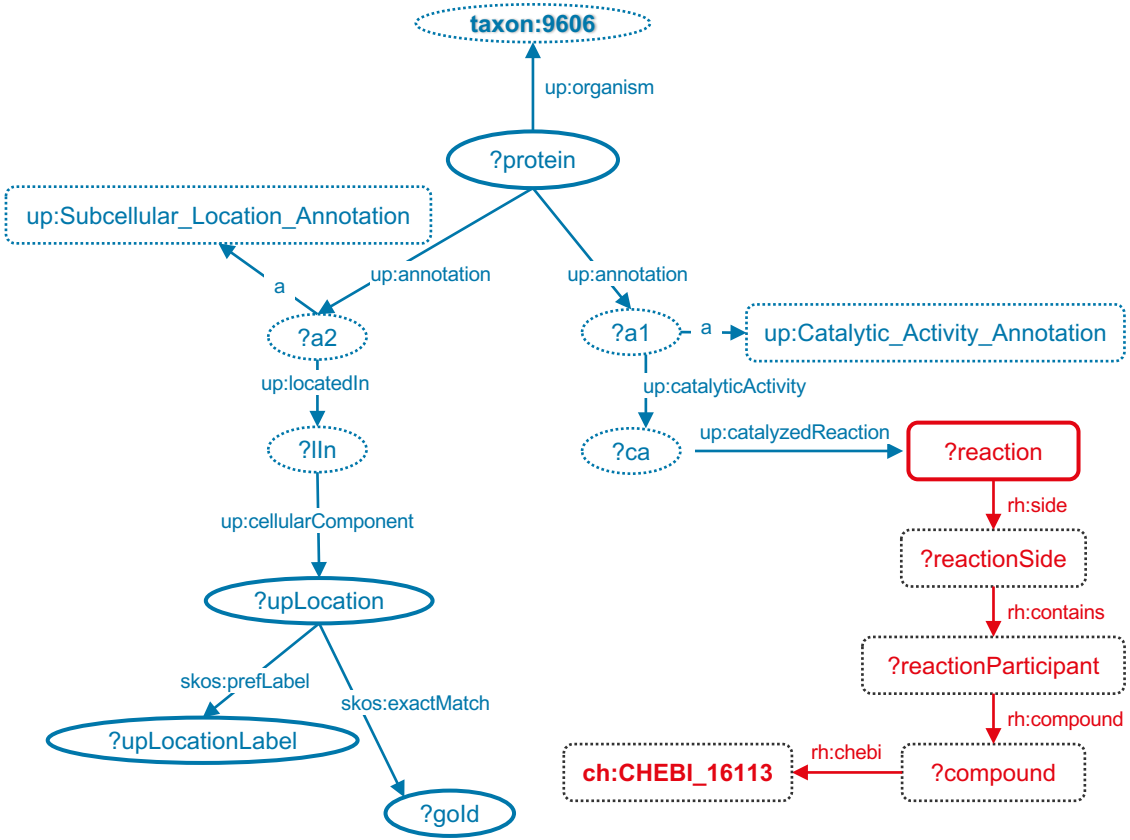
Expected: *false*

Answer (2019_10): *true* → to-do: fix errors

Query: where are the human enzymes metabolizing cholesterol **located in the cell?**



Q8: Where are the human enzymes metabolizing cholesterol **located in the cell**?




```
#endpoint: https://sparql.uniprot.org/sparql/
PREFIX rdfs:<http://www.w3.org/2000/01/rdf-schema#>
PREFIX rh:<http://rdf.rhea-db.org/>
PREFIX ch:<http://purl.obolibrary.org/obo/>
PREFIX up:<http://purl.uniprot.org/core/>
PREFIX taxon:http://purl.uniprot.org/taxonomy/
PREFIX skos:<http://www.w3.org/2004/02/skos/core#>
SELECT distinct ?protein ?reaction
                ?upLocation ?upLocationLabel ?goId
WHERE {
 SERVICE <https://sparql.rhea-db.org/sparql> {
  ?reaction rdfs:subClassOf rh:Reaction .
  ?reaction rh:equation ?reactionEquation .
  ?reaction rh:side ?reactionSide .
  ?reactionSide rh:contains ?participant .
  ?participant rh:compound ?compound .
  ?compound rh:chebi ch:CHEBI_16113 .
 }
 # Human proteins (taxid=9606)
 ?protein up:organism taxon:9606 .
 # Rhea catalayzed reactions
 ?protein up:annotation ?a1 .
 ?a1 a up:Catalytic_Activity_Annotation .
 ?a1 up:catalyticActivity ?ca .
 ?ca up:catalyzedReaction ?reaction .
 # UniProt cellular components
 ?protein up:annotation ?a2 .
 ?a2  a  up:Subcellular_Location_Annotation .
 ?a2 up:locatedIn ?IIn .
 ?IIn up:cellularComponent ?upLocation .
 ?upLocation skos:prefLabel ?upLocationLabel  .
 ?upLocation skos:exactMatch ?goId .
}
```

Q8: Where are the human enzymes metabolizing cholesterol **located in the cell**?

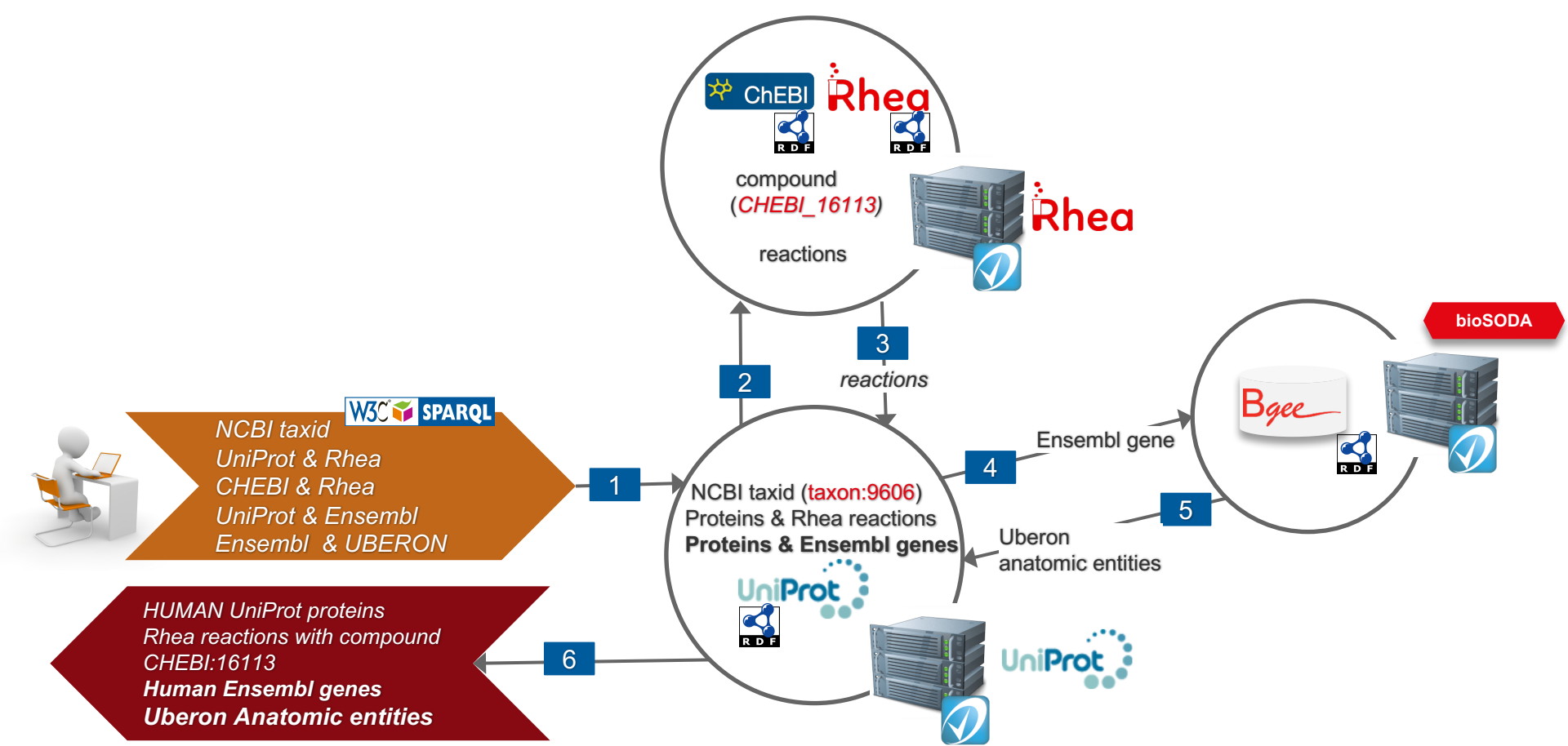
Results

Sparql XML | Sparql JSON | CSV | Show query | Share

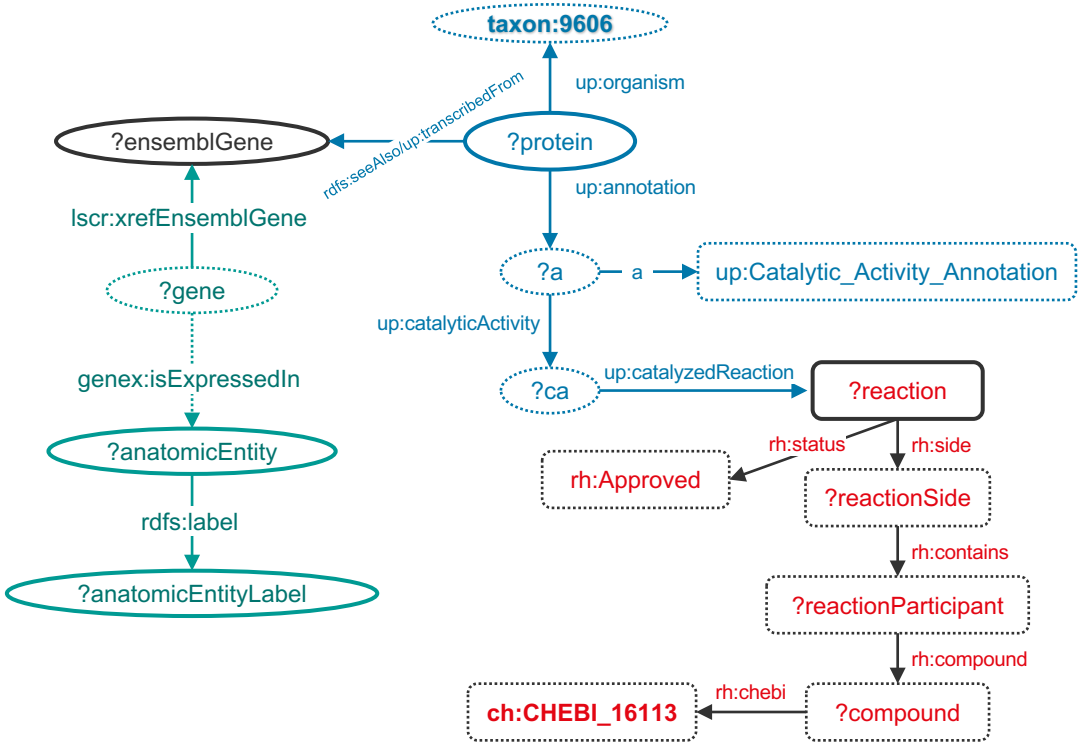
protein	proteinId	reaction	upLocation	upLocationLabel	gold
http://purl.uniprot.org/uniprot/O95477	"ABCA1_HUMAN"xsd:string	http://rdf.rhea-db.org/39051	http://purl.uniprot.org/locations/101	"Endosome"xsd:string	http://purl.obolibrary.org/obo/GO_0005768
http://purl.uniprot.org/uniprot/P45844	"ABCG1_HUMAN"xsd:string	http://rdf.rhea-db.org/39051	http://purl.uniprot.org/locations/134	"Golgi apparatus membrane"xsd:string	http://purl.obolibrary.org/obo/GO_0000139
http://purl.uniprot.org/uniprot/Q15392	"DHC24_HUMAN"xsd:string	http://rdf.rhea-db.org/36391	http://purl.uniprot.org/locations/134	"Golgi apparatus membrane"xsd:string	http://purl.obolibrary.org/obo/GO_0000139
http://purl.uniprot.org/uniprot/Q9HCG7	"GBA2_HUMAN"xsd:string	http://rdf.rhea-db.org/11956	http://purl.uniprot.org/locations/134	"Golgi apparatus membrane"xsd:string	http://purl.obolibrary.org/obo/GO_0000139
http://purl.uniprot.org/uniprot/Q9HCG7	"GBA2_HUMAN"xsd:string	http://rdf.rhea-db.org/58264	http://purl.uniprot.org/locations/134	"Golgi apparatus membrane"xsd:string	http://purl.obolibrary.org/obo/GO_0000139

[…]

Query: where are the human genes encoding enzymes metabolizing cholesterol **expressed**?



Q9: Where are the human genes encoding enzymes metabolizing cholesterol **expressed**?




```
#endpoint: https://sparql.uniprot.org/sparql/
PREFIX rdfs:<http://www.w3.org/2000/01/rdf-schema#>
PREFIX rh:<http://rdf.rhea-db.org/>
PREFIX ch:<http://purl.obolibrary.org/obo/>
PREFIX up:<http://purl.uniprot.org/core/>
PREFIX taxon:<http://purl.uniprot.org/taxonomy/>
PREFIX genex: <http://purl.org/genex#>
PREFIX lscr: <http://purl.org/lscr#>

SELECT distinct ?protein ?ensemblGene ?reaction
                ?anatomicEntityLabel ?anatomicEntity
WHERE {
 SERVICE <https://sparql.rhea-db.org/sparql> {
  ?reaction rdfs:subClassOf rh:Reaction .
  ?reaction rh:equation ?reactionEquation .
  ?reaction rh:side ?reactionSide .
  ?reactionSide rh:contains ?participant .
  ?participant rh:compound ?compound .
  ?compound rh:chebi ch:CHEBI_16113 .
 }

 ?protein up:organism taxon:9606 .
 ?protein up:annotation ?a .
 ?a a up:Catalytic_Activity_Annotation .
 ?a up:catalyticActivity ?ca .
 ?ca up:catalyzedReaction ?reaction .
 ?protein rdfs:seeAlso / up:transcribedFrom ?ensemblGene .

 SERVICE <http://biosoda.expasy.org/rdf4j-server/repositories/bgeelight> {
  ?gene genex:isExpressedIn ?anatomicEntity .
  ?gene lscr:xrefEnsemblGene ?ensemblGene .
  ?anatomicEntity rdfs:label ?anatomicEntityLabel .
 }
}
```

Q9: Where are the human genes encoding enzymes metabolizing cholesterol **expressed**?

UniProt

SPARQL Downloads Documentation/Help Contact

Results

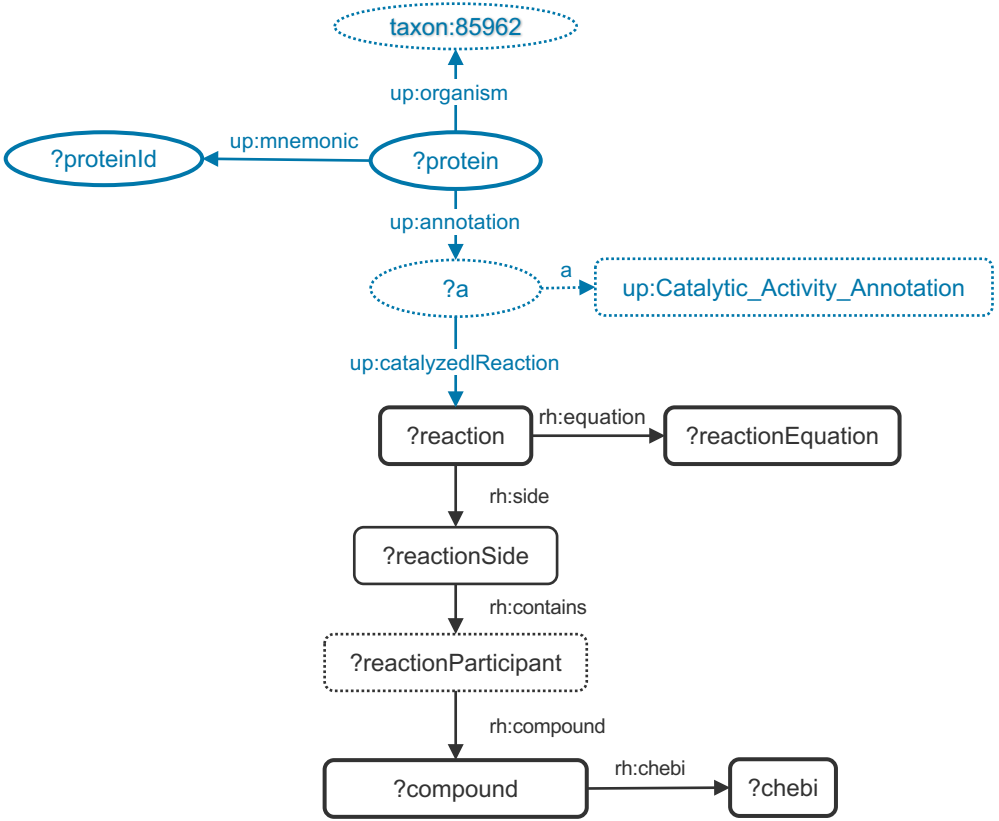
Sparql XML | Sparql JSON | CSV | Show query | Share

protein	ensemblGene	reaction	anatomicEntityLabel	anatomicEntity
uniprot:P08684	http://rdf.ebi.ac.uk/resource/ensembl/ENSG00000160868	rh:46140	"left testis"xsd:string	obo:UBERON_0004533
uniprot:P08684	http://rdf.ebi.ac.uk/resource/ensembl/ENSG00000160868	rh:46140	"pituitary gland"xsd:string	obo:UBERON_0000007
uniprot:P08684	http://rdf.ebi.ac.uk/resource/ensembl/ENSG00000160868	rh:46140	"anterior cingulate cortex"xsd:string	obo:UBERON_0009835
uniprot:P08684	http://rdf.ebi.ac.uk/resource/ensembl/ENSG00000160868	rh:46140	"ectocervix"xsd:string	obo:UBERON_0012249
uniprot:P08684	http://rdf.ebi.ac.uk/resource/ensembl/ENSG00000160868	rh:46140	"putamen"xsd:string	obo:UBERON_0001874
uniprot:P08684	http://rdf.ebi.ac.uk/resource/ensembl/ENSG00000160868	rh:46140	"right lung"xsd:string	obo:UBERON_0002167
uniprot:P08684	http://rdf.ebi.ac.uk/resource/ensembl/ENSG00000160868	rh:46140	"esophagogastric junction muscularis propria"xsd:string	obo:UBERON_0035841
uniprot:P08684	http://rdf.ebi.ac.uk/resource/ensembl/ENSG00000160868	rh:46140	"right ovary"xsd:string	obo:UBERON_0002118
uniprot:P08684	http://rdf.ebi.ac.uk/resource/ensembl/ENSG00000160868	rh:46140	"aorta"xsd:string	obo:UBERON_0000947
uniprot:P08684	http://rdf.ebi.ac.uk/resource/ensembl/ENSG00000160868	rh:46140	"minor salivary gland"xsd:string	obo:UBERON_0001830
uniprot:P08684	http://rdf.ebi.ac.uk/resource/ensembl/ENSG00000160868	rh:46140	"subcutaneous adipose tissue"xsd:string	obo:UBERON_0002190
uniprot:P08684	http://rdf.ebi.ac.uk/resource/ensembl/ENSG00000160868	rh:46140	"small intestine Peyer's patch"xsd:string	obo:UBERON_0003454
uniprot:P08684	http://rdf.ebi.ac.uk/resource/ensembl/ENSG00000160868	rh:46140	"blood"xsd:string	obo:UBERON_0000178

Explore/exploit the enormous richness of UniProt (Jerven)

- **Diseases**
- **3D structures (PDB)**
- **Homologs: orthoDB (Dima) and OMA (Tarcisio)**
- **Taxonomy / complete proteomes**
- **…**

Q10: Build the UniProt *H. pylori* proteome scale metabolic network




```
#endpoint: https://sparql.rhea-db.org/sparql

PREFIX rh:<http://rdf.rhea-db.org/>
PREFIX up:<http://purl.uniprot.org/core/>
PREFIX taxon:<http://purl.uniprot.org/taxonomy/>
SELECT  ?protein ?proteinId
        ?reaction ?reactionSide
        ?compound  ?chebi
        ?reactionEquation
WHERE {
 SERVICE <http://sparql.uniprot.org/sparql> {
  ?protein up:reviewed ?status .
  # Filter by NCBI taxid (H. pylori == 85962)
  ?protein up:organism taxon:85962 .
  ?protein up:mnemonic ?proteinId .
  # Rhea reactions catalyzed by UniProt proteins
  ?protein up:annotation ?a .
  ?a up:catalyticActivity ?ca .
  ?ca up:catalyzedReaction ?reaction .
 }
 ?reaction rh:equation ?reactionEquation .
 ?reaction rh:side ?reactionSide .
 ?reactionSide rh:contains ?participant .
 ?participant rh:compound ?compound .
 OPTIONAL {?compound rh:chebi ?chebi } .
}
```

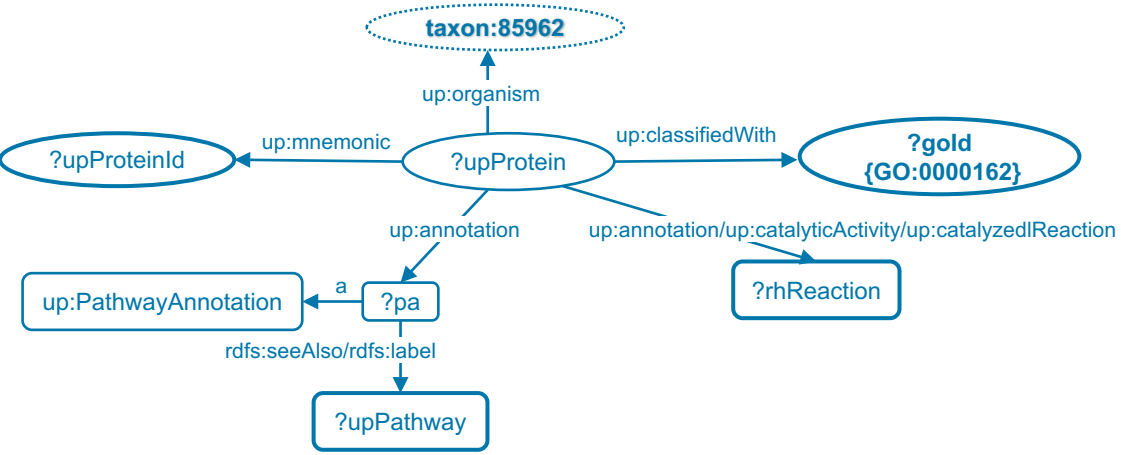
Q10: *H. pylori* metabolic network (UniProtKB/Swiss-Prot and Rhea-ChEBI)

protein	proteinId	reaction	reactionSide	compound	chebi	reactionEquation
http://purl.uniprot.org/uniprot/P56463	"DEOD_HELPY"	http://rdf.rhea-db.org/36431	http://rdf.rhea-db.org/36431_R	http://rdf.rhea-db.org/Compound_14976	http://purl.obolibrary.org/obo/CHEBI_26386	"a purine 2'-deoxy-D-ribonucleoside + phosphate = 2-deoxy-alpha-D-ribose 1-phosphate + a purine nucleobas
http://purl.uniprot.org/uniprot/P56463	"DEOD_HELPY"	http://rdf.rhea-db.org/36431	http://rdf.rhea-db.org/36431_L	http://rdf.rhea-db.org/Compound_15011	http://purl.obolibrary.org/obo/CHEBI_142361	"a purine 2'-deoxy-D-ribonucleoside + phosphate = 2-deoxy-alpha-D-ribose 1-phosphate + a purine nucleobas
http://purl.uniprot.org/uniprot/O26027	"MURE_HELPY"	http://rdf.rhea-db.org/23676	http://rdf.rhea-db.org/23676_R	http://rdf.rhea-db.org/Compound_10767	http://purl.obolibrary.org/obo/CHEBI_83905	"ATP + meso-2,6-diaminopimelate + UDP-N-acetyl-alpha-D-muramoyl-L-alanyl-D-glutamate = ADP + H(+) + phosp
http://purl.uniprot.org/uniprot/O25016	"KDSB_HELPY"	http://rdf.rhea-db.org/23448	http://rdf.rhea-db.org/23448_R	http://rdf.rhea-db.org/Compound_11525	http://purl.obolibrary.org/obo/CHEBI_85987	"3-deoxy-alpha-D-manno-oct-2-ulosonate + CTP = CMP-3-deoxy-beta-D-manno-octulosonate + diphosphate"
http://purl.uniprot.org/uniprot/O25611	"O25611_HELPY"	http://rdf.rhea-db.org/28066	http://rdf.rhea-db.org/28066_L	http://rdf.rhea-db.org/Compound_11525	http://purl.obolibrary.org/obo/CHEBI_85987	"CMP-3-deoxy-beta-D-manno-octulosonate + lipid IVA (E. coli) = alpha-Kdo-(2->6)-lipid IVA + CMP + H(+)"
http://purl.uniprot.org/uniprot/P56006	"SCOA_HELPY"	http://rdf.rhea-db.org/24564	http://rdf.rhea-db.org/24564_R	http://rdf.rhea-db.org/Compound_12177	http://purl.obolibrary.org/obo/CHEBI_90726	"a 3-oxo acid + succinyl-CoA = a 3-oxoacyl-CoA + succinate"
http://purl.uniprot.org/uniprot/P56007	"SCOB_HELPY"	http://rdf.rhea-db.org/24564	http://rdf.rhea-db.org/24564_R	http://rdf.rhea-db.org/Compound_12177	http://purl.obolibrary.org/obo/CHEBI_90726	"a 3-oxo acid + succinyl-CoA = a 3-oxoacyl-CoA + succinate"
http://purl.uniprot.org/uniprot/P56122	"AROC_HELPY"	http://rdf.rhea-db.org/21020	http://rdf.rhea-db.org/21020_L	http://rdf.rhea-db.org/Compound_1720	http://purl.obolibrary.org/obo/CHEBI_57701	"5-O-(1-carboxyvinyl)-3-phosphoshikimate = chorismate + phosphate"
http://purl.uniprot.org/uniprot/P56197	"AROA_HELPY"	http://rdf.rhea-db.org/21256	http://rdf.rhea-db.org/21256_R	http://rdf.rhea-db.org/Compound_1720	http://purl.obolibrary.org/obo/CHEBI_57701	"3-phosphoshikimate + phosphoenolpyruvate = 5-O-(1-carboxyvinyl)-3-phosphoshikimate + phosphate"
http://purl.uniprot.org/uniprot/O24865	"O24865_HELPY"	http://rdf.rhea-db.org/34099	http://rdf.rhea-db.org/34099_R	http://rdf.rhea-db.org/Compound_2551	http://purl.obolibrary.org/obo/CHEBI_57920	"carboxynorspermidine + H(+) = CO2 + norspermidine"

[…]

→ To compare with published GSMNs (MetaNetX, Marco)

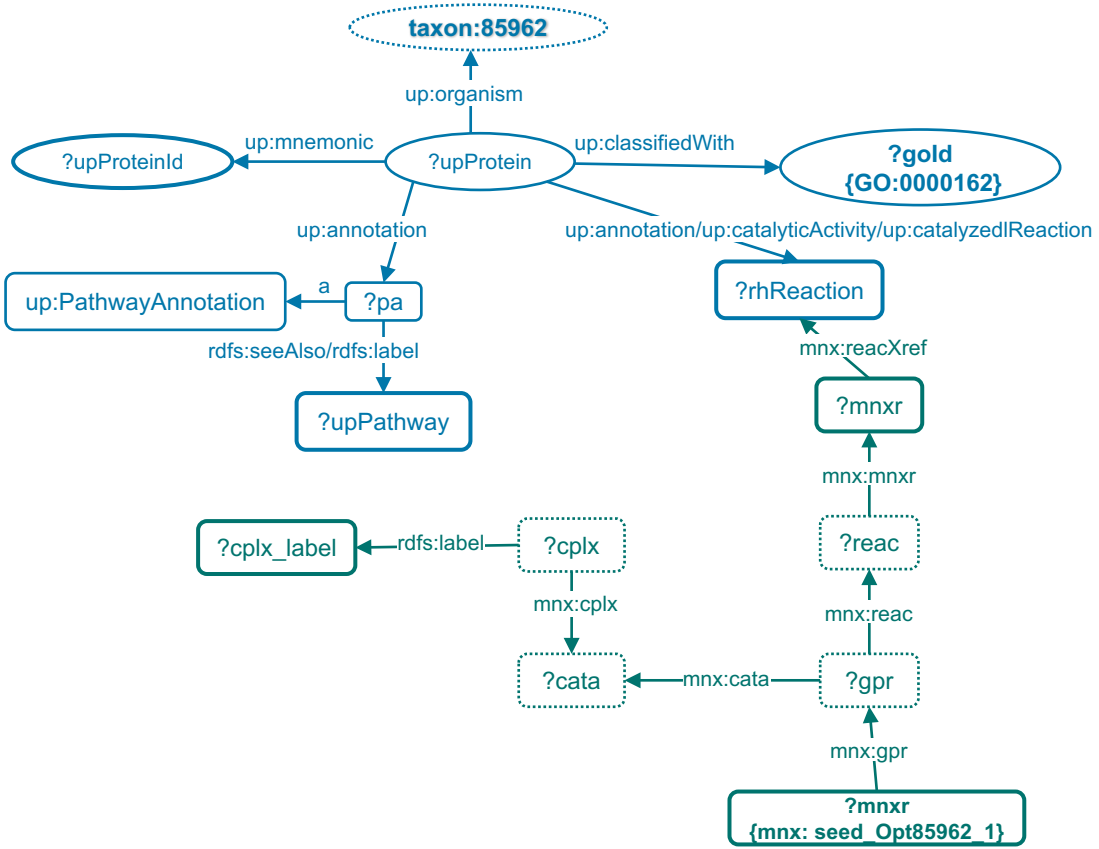
Q11: Explore *H. pylori* tryptophan biosynthesis pathway (GO:0000162)




```
#endpoint: https://sparql.uniprot.org/sparql/
PREFIX rdfs:<http://www.w3.org/2000/01/rdf-schema#>
PREFIX up:<http://purl.uniprot.org/core/>
PREFIX rh:<http://rdf.rhea-db.org/>
PREFIX taxon:<http://purl.uniprot.org/taxonomy/>
PREFIX ch:<http://purl.obolibrary.org/obo/>
PREFIX GO:<http://purl.obolibrary.org/obo/GO_>
SELECT distinct ?upProteinId ?gold ?rhReaction ?upPathway
WHERE{
 ?upProtein up:reviewed true .
 ?upProtein up:mnemonic  ?upProteinId .
 # HELPY proteins (taxid=85962)
 ?upProtein up:organism taxon:85962 .
 ?upProtein up:annotation/up:catalyticActivity/up:catalyzedReaction ?rhReaction .
 ?upProtein up:classifiedWith ?gold .
 VALUES ?gold {GO:0000162} # GO: tryptophan biosynthesis
 OPTIONAL {?upProtein up:annotation ?pa . # Pathway annotation
           ?pa a up:Pathway_Annotation .
           ?pa rdfs:seeAlso/rdfs:label ?upPathway . }
}
ORDER BY ?upPathway
```

upProteinId	gold	rhReaction	upPathway
"TRPE_HELPY"xsd:string	http://purl.obolibrary.org/obo/GO_0000162	http://rdf.rhea-db.org/21732	"Amino-acid biosynthesis; L-tryptophan biosynthesis; L-tryptophan from chorismate: step 1/5"xsd:string
"TRPG_HELPY"xsd:string	http://purl.obolibrary.org/obo/GO_0000162	http://rdf.rhea-db.org/21732	"Amino-acid biosynthesis; L-tryptophan biosynthesis; L-tryptophan from chorismate: step 1/5"xsd:string
"TRPD_HELPY"xsd:string	http://purl.obolibrary.org/obo/GO_0000162	http://rdf.rhea-db.org/11768	"Amino-acid biosynthesis; L-tryptophan biosynthesis; L-tryptophan from chorismate: step 2/5"xsd:string
"TRPC_HELPY"xsd:string	http://purl.obolibrary.org/obo/GO_0000162	http://rdf.rhea-db.org/23476	"Amino-acid biosynthesis; L-tryptophan biosynthesis; L-tryptophan from chorismate: step 3/5"xsd:string
"TRPC_HELPY"xsd:string	http://purl.obolibrary.org/obo/GO_0000162	http://rdf.rhea-db.org/21540	"Amino-acid biosynthesis; L-tryptophan biosynthesis; L-tryptophan from chorismate: step 3/5"xsd:string
"TRPC_HELPY"xsd:string	http://purl.obolibrary.org/obo/GO_0000162	http://rdf.rhea-db.org/23476	"Amino-acid biosynthesis; L-tryptophan biosynthesis; L-tryptophan from chorismate: step 4/5"xsd:string
"TRPC_HELPY"xsd:string	http://purl.obolibrary.org/obo/GO_0000162	http://rdf.rhea-db.org/21540	"Amino-acid biosynthesis; L-tryptophan biosynthesis; L-tryptophan from chorismate: step 4/5"xsd:string
"TRPA_HELPY"xsd:string	http://purl.obolibrary.org/obo/GO_0000162	http://rdf.rhea-db.org/10532	"Amino-acid biosynthesis; L-tryptophan biosynthesis; L-tryptophan from chorismate: step 5/5"xsd:string
"TRPB_HELPY"xsd:string	http://purl.obolibrary.org/obo/GO_0000162	http://rdf.rhea-db.org/10532	"Amino-acid biosynthesis; L-tryptophan biosynthesis; L-tryptophan from chorismate: step 5/5"xsd:string

Q12: H. pylori enzyme complexes for tryptophan biosynthesis pathway (MetaNetX: seed_Opt85962_1)




```
#endpoint: https://sparql.uniprot.org/sparql/
PREFIX rdfs:<http://www.w3.org/2000/01/rdf-schema#>
PREFIX up:<http://purl.uniprot.org/core/>
PREFIX rh:<http://rdf.rhea-db.org/>
PREFIX taxon:<http://purl.uniprot.org/taxonomy/>
PREFIX mnx:<https://rdf.metanetx.org/schema/>
PREFIX mnet:<https://rdf.metanetx.org/mnet/>
PREFIX GO:<http://purl.obolibrary.org/obo/GO_>

SELECT distinct ?upPathway ?upProteinId ?rhReaction
                ?mnxr ?cplx_label ?mnet
WHERE{
   ?upProtein up:reviewed true .
   ?upProtein up:mnemonic ?upProteinId .
   ?upProtein up:organism taxon:85962 .
   ?upProtein up:classifiedWith GO:0000162.
   ?upProtein up:annotation/up:catalyticActivity/up:catalyzedReaction ?rhReaction
.
   # Pathway annotation
   OPTIONAL {?upProtein up:annotation ?pa          .
          ?pa a up:Pathway_Annotation          ;
             rdfs:seeAlso/rdfs:label ?upPathway . }
   SERVICE <https://rdf.metanetx.org/sparql> {
      ?mnxr mnx:reacXref ?rhReaction .
      ?reac mnx:mnxr ?mnxr .
      ?gpr  mnx:reac ?reac .
      ?gpr  mnx:cata ?cata .
      ?cata mnx:cplx ?cplx .
      ?cplx rdfs:label ?cplx_label .
      ?mnet mnx:gpr ?gpr .
      VALUES ?mnet {mnet:seed_Opt85962_1}
   }
}
ORDER BY ?upPathway
```

Q12: H. pylori enzyme complexes for tryptophan biosynthesis pathway (MetaNetX: seed_Opt85962_1)

upPathway	upProteinId	rhReaction	mnxr	cplx_label	mnet
"Amino-acid biosynthesis; L-tryptophan biosynthesis; L-tryptophan from chorismate: step 1/5"xsd:string	"TRPE_HELPY"xsd:string	http://rdf.rhea-db.org/21732	https://rdf.metanetx.org/reac/MNXR95843	"bact:TRPE_HELPY+bact:TRPG_HELPY"xsd:string	https://rdf.metanetx.org/mnet/seed_Opt85962_1
"Amino-acid biosynthesis; L-tryptophan biosynthesis; L-tryptophan from chorismate: step 1/5"xsd:string	"TRPG_HELPY"xsd:string	http://rdf.rhea-db.org/21732	https://rdf.metanetx.org/reac/MNXR95843	"bact:TRPE_HELPY+bact:TRPG_HELPY"xsd:string	https://rdf.metanetx.org/mnet/seed_Opt85962_1
"Amino-acid biosynthesis; L-tryptophan biosynthesis; L-tryptophan from chorismate: step 2/5"xsd:string	"TRPD_HELPY"xsd:string	http://rdf.rhea-db.org/11768	https://rdf.metanetx.org/reac/MNXR95842	"bact:TRPD_HELPY"xsd:string	https://rdf.metanetx.org/mnet/seed_Opt85962_1
"Amino-acid biosynthesis; L-tryptophan biosynthesis; L-tryptophan from chorismate: step 3/5"xsd:string	"TRPC_HELPY"xsd:string	http://rdf.rhea-db.org/23476	https://rdf.metanetx.org/reac/MNXR100814	"bact:TRPC_HELPY"xsd:string	https://rdf.metanetx.org/mnet/seed_Opt85962_1
"Amino-acid biosynthesis; L-tryptophan biosynthesis; L-tryptophan from chorismate: step 3/5"xsd:string	"TRPC_HELPY"xsd:string	http://rdf.rhea-db.org/21540	https://rdf.metanetx.org/reac/MNXR103156	"bact:TRPC_HELPY"xsd:string	https://rdf.metanetx.org/mnet/seed_Opt85962_1
"Amino-acid biosynthesis; L-tryptophan biosynthesis; L-tryptophan from chorismate: step 4/5"xsd:string	"TRPC_HELPY"xsd:string	http://rdf.rhea-db.org/21540	https://rdf.metanetx.org/reac/MNXR103156	"bact:TRPC_HELPY"xsd:string	https://rdf.metanetx.org/mnet/seed_Opt85962_1
"Amino-acid biosynthesis; L-tryptophan biosynthesis; L-tryptophan from chorismate: step 4/5"xsd:string	"TRPC_HELPY"xsd:string	http://rdf.rhea-db.org/23476	https://rdf.metanetx.org/reac/MNXR100814	"bact:TRPC_HELPY"xsd:string	https://rdf.metanetx.org/mnet/seed_Opt85962_1
"Amino-acid biosynthesis; L-tryptophan biosynthesis; L-tryptophan from chorismate: step 5/5"xsd:string	"TRPB_HELPY"xsd:string	http://rdf.rhea-db.org/10532	https://rdf.metanetx.org/reac/MNXR104343	"bact:TRPA_HELPY+bact:TRPB_HELPY"xsd:string	https://rdf.metanetx.org/mnet/seed_Opt85962_1
"Amino-acid biosynthesis; L-tryptophan biosynthesis; L-tryptophan from chorismate: step 5/5"xsd:string	"TRPA_HELPY"xsd:string	http://rdf.rhea-db.org/10532	https://rdf.metanetx.org/reac/MNXR104343	"bact:TRPA_HELPY+bact:TRPB_HELPY"xsd:string	https://rdf.metanetx.org/mnet/seed_Opt85962_1

Programmatic access

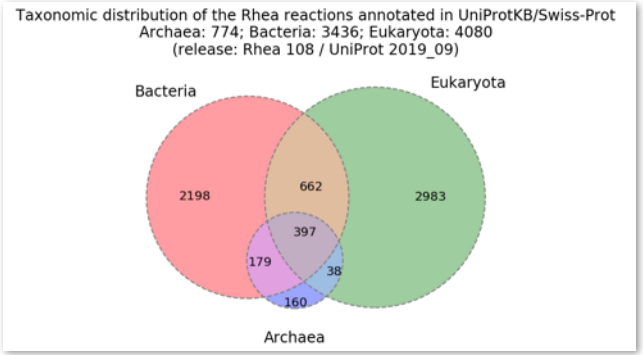
jupyter

SPARQL

+

python

SPARQLWrapper
pandas
matplotlib
matplotlib_venn
[...]



Rhea reactions used in UniProtKB/Swiss-Prot annotation (release: Rhea 108 / UniProt 2019_09)

		label	count	pct
category				
domain	archaea	857	8.6%
domain	bacteria	3334	33.6%
domain	eukaryota	4221	42.5%
domain	viruses	187	1.9%

Jupyter notebook: rhea_tutorial_SWAT4HCLS_2019.ipynb

2.1 Python Libraries

```
from SPARQLWrapper import SPARQLWrapper, JSON
import pandas as pd
from pandas.io.json import json_normalize
```

2.2 SPARQL endpoints

```
sparql_uniprot_url = "https://sparql.uniprot.org/sparql/"
sparql_rhea_url = "https://sparql.rhea-db.org/sparql"
```

2.3 Python functions

```
def sparql2pandas(sparql_query, sparql_service_url):
    """
    Use SPARQLWrapper to query a SPARQL endpoint (sparql_service_url) with a given SPARQL query string
    (sparql_query) that uses SELECT query form.
    Return the SPARQL query result as a pandas Dataframe.
    """
    #
    # run the SPARQL query
    #
    sparql=SPARQLWrapper(sparql_service_url)
    sparql.setQuery(sparql_query)
    sparql.setReturnFormat(JSON)
    res = sparql.query().convert()
```

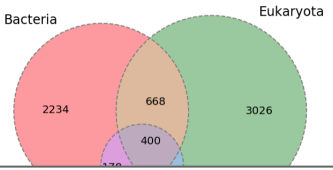
3.1 Q1: Retrieve all Rhea reactions (approved or preliminary) and their chemical equation

```
sparql_Q1="""
#endpoint: https://sparql.rhea-db.org/sparql

PREFIX rh:<http://rdf.rhea-db.org/>

SELECT ?reaction ?reactionEquation
WHERE {
  ?reaction rdfs:subClassOf rh:Reaction .
  ?reaction rh:status ?status .
  VALUES ?status {rh:Approved rh:Preliminary}
  ?reaction rh:equation ?reactionEquation .
}
ORDER BY ?reaction
"""
print(sparql_Q1)
```

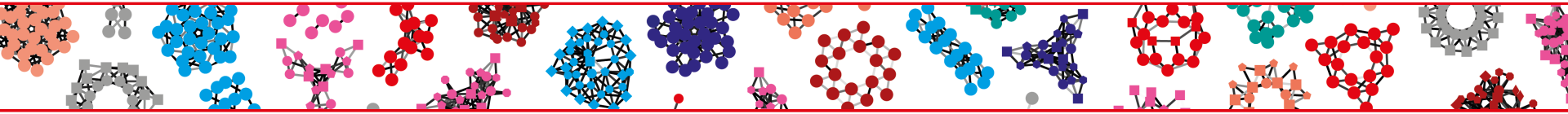
4.1.2 Compute Venn diagram using matplotlib and matplotlib_venn python libraries

```
import matplotlib.pyplot as plt
from matplotlib_venn import venn3
from matplotlib_venn import venn3_circles

set_A = set(df_up_rh[df_up_rh.domainName == 'Archaea']['rhReaction'].drop_duplicates())
set_B = set(df_up_rh[df_up_rh.domainName == 'Bacteria']['rhReaction'].drop_duplicates())
set_E = set(df_up_rh[df_up_rh.domainName == 'Eukaryota']['rhReaction'].drop_duplicates())

nb_A = len(set_A)
nb_B = len(set_B)
nb_E = len(set_E)
nb_all = len(set_A | set_B | set_E)

title = 'Taxonomic distribution of the Rhea reactions annotated in UniProtKB/Swiss-Prot\n'
title += 'Rhea reactions: in Archaea=%s ; in Bacteria=%s; in Eukaryota=%s\n' % (nb_A, nb_B, nb_E)

# Plot Venn diagram
v=venn3([set_B,set_E,set_A], set_labels = ('Bacteria','Eukaryota', '\nArchaea'))
c=venn3_circles([set_B,set_E,set_A], linestyle='dashed', linewidth=1, color="grey")
plt.title(title)

plt.show();
```

Taxonomic distribution of the Rhea reactions annotated in UniProtKB/Swiss-Prot
Rhea reactions: in Archaea=777 ; in Bacteria=3480; in Eukaryota=4132



Overview



Summary

- **How to query metabolic data across multiple data source :**

Source	Description	web site	SPAQRL endpoint
Rhea	chemical reactions	https://www.rhea-db.org/	https://sparql.uniprot.org/sparql/
UniProt	protein sequences and annotations	https://www.uniprot.org/	https://sparql.rhea-db.org/sparql
Bgee	expression data	https://bgee.org/	hosted by http://biosoda.expasy.org/
MetaNetX	reconcilied metabolic networks	https://www.metanetx.org/	https://rdf.metanetx.org/

  - Do simple chemical and reaction searches
  - Do structural search using InChiKey
  - How to use ASK for QC
  - How to link chemical data to enzymes, subcellular location or anatomic entities
  - How to enter in the world of metabolic models

- **Programmatic access in Python**

- **Conclusion: very powerful** but... still some technical issues to be resolved

Acknowledgements

Photograph taken by Chris James

Swiss-Prot group director
Alan Bridge

Biocurators
Lucila Aimo, Kristian Axelsen, Elisabeth Coudert, **Anne Morgat**, Nevila Nouspikel

Developers
Thierry Lombardot (Rhea SPARQL), Jerven Bolleman (UniProt SPARQL), Teresa Neto (UniProt/Rhea/ChEBI synchronization), Sébastien Géhant (discussions), **Nicole Redaschi** (head)

Rhea development and curation activities at the SIB are supported by the Swiss Federal Government through the State Secretariat for Education, Research and Innovation (SERI).

Rhea SPARQL endpoint development was supported in part by **ELIXIR** (Implementation study on "A microbial metabolism resource for Systems Biology").

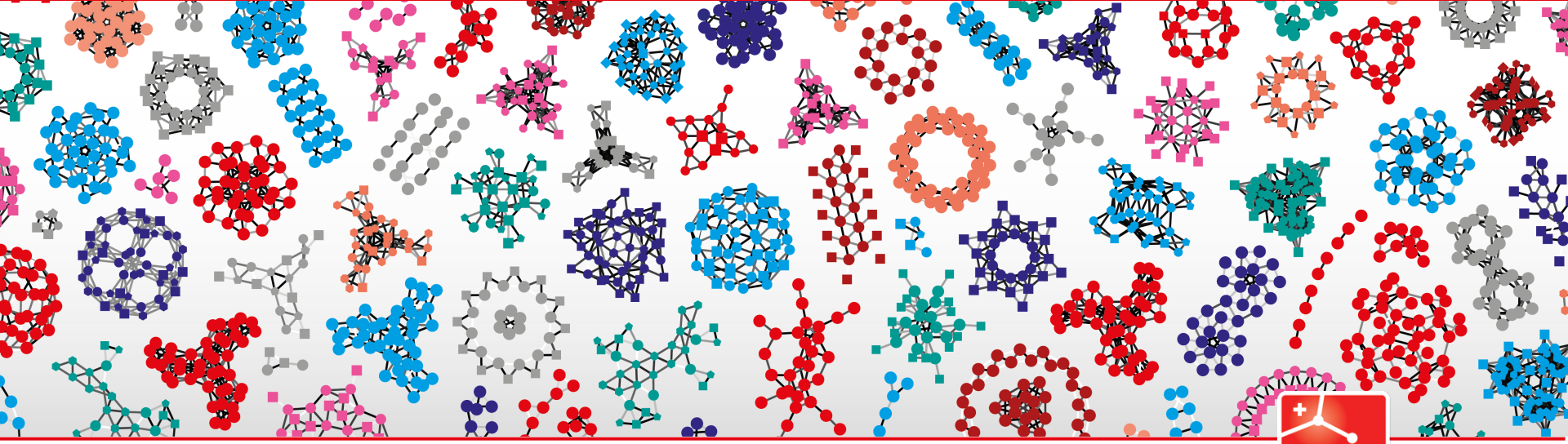
We gratefully acknowledge the software contributions of **ChemAxon**.

Web: https://www.rhea-db.org/

SPARQL: https://sparql.rhea-db.org/sparql

Email support: rhea-reaction@sib.swiss

Twitter: @rhea_db



SIB
Swiss Institute of Bioinformatics

Thank you for your attention

anne.morgat@sib.swiss