SPARQLing Rhea

Anne Morgat, Swiss-Prot group
Accompanying Jupyter notebook:
Overview

01 What is Rhea?
02 Data model
03 SPARQLing Rhea
04 Summary
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


formula: C23 H30 N8 O9, charge: -2

formula: C23 H30 N8 O9, charge: -2
5,10-methylenetetrahydrofolate + D-alanine + H₂O = tetrahydrofolate + 2-methylserine
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 1R-hydroxy group of β-D-glucose with the 3R-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.

**Chemical Properties**
- **Formula**: C30H49O9
- **Net Charge**: -1
- **Average Mass**: 553.706
- **Monoisotopic Mass**: 553.33821
- **InChI**: InChI=1S/C30H49O9/c1-15(4-7-23(33)34)18-5-6-19-24-20(9-11-30(18,19)3)29(10-8-17(12-16(29)13-21(24)32)38-29-37(26)36(35)32)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,-23,+25,-26,+27,+28,+29,-30,-31/m1/s1
- **InChIKey**: QRLLJGDVRXVH-QD-UMPBQQUS-A
- **SMILES**: C1[C@@]2[[C@H]3[[C@H][C@H][C@H][C@H][C@H][C@H][C@H][C@H][C@H][C@H][C@H][C@H][C@H][C@H][C@H][C@H][C@H][C@H][C@H][C@H][C@H][C@H][C@H][C@H][C@H][C@H][C@H][C@H][C@H][C@H][C@H][C@H][C@H][C@H][C@H][C@H][C@H][C@H][C@H][C@H][C@H][C@H][C@H][C@H][C@H][C@H][C@H][C@H][C@H][C@H][C@H][C@H][C@H][C@H][C@H][C@H][C@H][C@H][C@H][C@H][C@H][C@H][C@H][C@H][C@H][C@H][C@H][C@H]
ChEBI:142610 - relationships
**CHEBI:142610 - chenodeoxycholate-3-O-β-D-glucoside(1-)**

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

---

**RHEA:58340**

- **betaglucosyl-(1->3)-O-chenodeoxychololate**
  - CHEBI:142610
  - Formula: C30H49O9
  - Charge: -1

- **H2O**
  - Formula: H2O
  - Charge: 0

- **chenodeoxycholate**
  - CHEBI:36234
  - Formula: C24H39O4
  - Charge: -1

- **D-glucose**
  - CHEBI:4167
  - Formula: C6H12O6
  - Charge: 0
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 curation: choice of reaction participants

5,10-methylenetetrahydrofolate + D-alanine + H₂O = tetrahydrofolate + 2-methylserine
Different chemical species for the same compound

ChEBI relationships:

- Is conjugate acid of
- Is conjugate base of
- Is tautomer of

Diagram showing relationships between D-alanine, D-alanine zwitterion, D-alaninate, D-alaninium, and their respective CHEBI identifiers.
Rhea curation: choice of reaction participants

5,10-methylenetetrahydrofolate + D-alanine + H₂O = tetrahydrofolate + 2-methylserine

Major microspecies at pH 7.3

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

Protein

Nucleic acid

10746 Rhea reaction participants (Rel 109)

84.5%

SmallMolecule (9081)

GenericPeptide (898)

GenericPolynucleotide (540)

Polymer (227)

2.1%

5.0%

8.4%
Rhea reactions hierarchical classification (exploit ChEBI ontology)
RHEA:21816 (APPROVED)
a D-amino acid + H₂O + O₂ = a 2-oxo carboxylate + H₂O₂ + NH₄⁺(+)


RHEA:22688 (APPROVED)
D-alanine + H₂O + O₂ = H₂O₂ + NH₄⁺(+) + pyruvate

RHEA:21816  (APPROVED)
a D-amino acid + H2O + O2 = a 2-oxo carboxylate + H2O2 + NH4(+)


RHEA:22688  (APPROVED)
D-alanine + H2O + O2 = H2O2 + NH4(+) + pyruvate

EC 1.4.3.3 (D-amino-acid oxidase)

RHEA:21816 (APPROVED)

\[ \text{a D-amino acid} + \text{H}_2\text{O} + \text{O}_2 \rightarrow \text{a 2-oxo carboxylate} + \text{H}_2\text{O}_2 + \text{NH}_4^+ \]

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

EC 1.4.3.19 (alternate reaction of glycine oxidase)

RHEA:22688 (APPROVED)

\[ \text{D-alanine} + \text{H}_2\text{O} + \text{O}_2 \rightarrow \text{H}_2\text{O}_2 + \text{NH}_4^+ + \text{pyruvate} \]
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+

• EC 2 Transferases

• EC 3 Hydrolases

• EC 4 Lyases

• EC 5 Isomerases

• EC 6 Ligases

• EC 7 Translocases

(is a)

missing deeper levels

is a
Rhea and Enzyme classification

Rhea and IUBMB enzyme classification
12290 reactions
(Rhea release: 109)

5594 reactions linked to EC numbers
(45.5%)

6696 reactions not linked to EC numbers
(54.5%)

6410 EC numbers in IUBMB enzyme classification
(releases UniProt 2019_10 / Rhea 109)

5645 EC numbers linked to reactions
(88.1%)

765 EC numbers NOT linked to reactions
= IUBMB text reaction
(11.9%)
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%)
Overview

01. What is Rhea?

02. Data model

03. SPARQLing Rhea

04. Summary
Data model (https://www.rhea-db.org/rhea_rdf_documentation.pdf)
Reactions, sides and participants

RHEA:10064 (APPROVED)

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

Reactions participants: rh:Compound classes

Diagram showing the categorization of reaction participants into rh:SmallMolecule, rh:GenericCompound, and rh:Polymer, with further subclassifications into rh:GenericPolypeptide, rh:GenericPolynucleotide, and rh:GenericHeteropolysaccharide.
Reactions participants: links to ChEBI
Rhea SPARQL endpoint: rhea.rdf + chebi.owl

https://sparql.rhea-db.org/sparql
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 .
  VALUES ?status {rh:Approved rh:Preliminary}
}
ORDER BY ?reaction
```

<table>
<thead>
<tr>
<th>reaction</th>
<th>reactionEquation</th>
</tr>
</thead>
<tbody>
<tr>
<td><a href="http://rdf.rhea-db.org/10000">http://rdf.rhea-db.org/10000</a></td>
<td>&quot;H2O + pentamidine = NH4(+) + pentanoate&quot;</td>
</tr>
<tr>
<td><a href="http://rdf.rhea-db.org/10004">http://rdf.rhea-db.org/10004</a></td>
<td>&quot;benzyl isothiocyanate = benzyl thiocyanate&quot;</td>
</tr>
<tr>
<td><a href="http://rdf.rhea-db.org/10008">http://rdf.rhea-db.org/10008</a></td>
<td>&quot;[protein]-dithiol + a hydroperoxide = [protein]-disulfide + an alcohol + H2O&quot;</td>
</tr>
<tr>
<td><a href="http://rdf.rhea-db.org/10012">http://rdf.rhea-db.org/10012</a></td>
<td>&quot;(R)-6-hydroxynicotine + H2O + O2 = 6-hydroxy-pseudo-6-hydroxynicotine + H2O2&quot;</td>
</tr>
<tr>
<td><a href="http://rdf.rhea-db.org/10016">http://rdf.rhea-db.org/10016</a></td>
<td>&quot;H2O + O-sinapoylcholine = choline + H(+) + trans-sinapate&quot;</td>
</tr>
<tr>
<td><a href="http://rdf.rhea-db.org/10020">http://rdf.rhea-db.org/10020</a></td>
<td>&quot;H2O + L-saccaropine + NADP(+) = (S)-2-amino-6-oxohexanoate + H(+) + L-glutamate + NADPH&quot;</td>
</tr>
<tr>
<td><a href="http://rdf.rhea-db.org/10028">http://rdf.rhea-db.org/10028</a></td>
<td>&quot;D-glutamate + H2O + O2 = 2-oxoglutarate + H2O2 + NH4(+)&quot;</td>
</tr>
<tr>
<td><a href="http://rdf.rhea-db.org/10032">http://rdf.rhea-db.org/10032</a></td>
<td>&quot;[protein]-C-terminal-L-glutamine + H2O = [protein]-C-terminal-L-glutamate + NH4(+)&quot;</td>
</tr>
<tr>
<td><a href="http://rdf.rhea-db.org/10036">http://rdf.rhea-db.org/10036</a></td>
<td>&quot;4-hydroxy-3-methoxybenzenemethanol + O2 = 4-hydroxy-3-methoxybenzaldehyde + H2O2&quot;</td>
</tr>
</tbody>
</table>

[...]
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

#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 .
  ?compound1 rh:chebi ch:CHEBI_29985 .
  ?reactionSide1 rh:transformableTo ?reactionSide2 .
}

CHEBI:29985 (L-glutamate)  CHEBI:58359 (L-glutamine)
Q3: Select all approved reactions with CHEBI:17815 (a 1,2-diacyl-sn-glycerol) or one of its descendant.

```sparql
#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 .
  ?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.

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

[...]
**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])[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

```sparql
#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 .  
  VALUES (?status) {(rh:Approved) (rh:Preliminary)}  
  ?chebi up:name ?chebiUniprotName .  
  VALUES (?inchikey) {("HVYWMOMLDIMFJA-DPAQBDIFSA-N")}  
  ?chebiUniprotName up:name ?chebiUniprotName .  
}  
ORDER BY ?reaction
```
Q4: Retrieve Rhea reactions that involve cholesterol using its InChiKey

<table>
<thead>
<tr>
<th>chebi</th>
<th>chebiUnprotName</th>
<th>reaction</th>
<th>reactionEquation</th>
</tr>
</thead>
<tbody>
<tr>
<td><a href="http://purl.obolibrary.org/obo/CHEBI_16113">http://purl.obolibrary.org/obo/CHEBI_16113</a></td>
<td>&quot;cholesterol&quot;</td>
<td><a href="http://rdf.rhea-db.org/11956">http://rdf.rhea-db.org/11956</a></td>
<td>&quot;cholesteryl-beta-D-glucoside + H2O = cholesterol + D-glucose&quot;</td>
</tr>
<tr>
<td><a href="http://purl.obolibrary.org/obo/CHEBI_16113">http://purl.obolibrary.org/obo/CHEBI_16113</a></td>
<td>&quot;cholesterol&quot;</td>
<td><a href="http://rdf.rhea-db.org/17729">http://rdf.rhea-db.org/17729</a></td>
<td>&quot;en acyl-CoA + cholesterol = a cholesterol ester + CoA&quot;</td>
</tr>
<tr>
<td><a href="http://purl.obolibrary.org/obo/CHEBI_16113">http://purl.obolibrary.org/obo/CHEBI_16113</a></td>
<td>&quot;cholesterol&quot;</td>
<td><a href="http://rdf.rhea-db.org/21328">http://rdf.rhea-db.org/21328</a></td>
<td>&quot;cholesterol + O2 = chol-4-en-3-one + H2O2&quot;</td>
</tr>
<tr>
<td><a href="http://purl.obolibrary.org/obo/CHEBI_16113">http://purl.obolibrary.org/obo/CHEBI_16113</a></td>
<td>&quot;cholesterol&quot;</td>
<td><a href="http://rdf.rhea-db.org/23984">http://rdf.rhea-db.org/23984</a></td>
<td>&quot;cholesterol + NADP(+) = cholest-5,7-dien-3beta-ol + H(+) + NADPH&quot;</td>
</tr>
<tr>
<td><a href="http://purl.obolibrary.org/obo/CHEBI_16113">http://purl.obolibrary.org/obo/CHEBI_16113</a></td>
<td>&quot;cholesterol&quot;</td>
<td><a href="http://rdf.rhea-db.org/32183">http://rdf.rhea-db.org/32183</a></td>
<td>&quot;cholesterol + O2 = chol-5-en-3-one + H2O2&quot;</td>
</tr>
<tr>
<td><a href="http://purl.obolibrary.org/obo/CHEBI_16113">http://purl.obolibrary.org/obo/CHEBI_16113</a></td>
<td>&quot;cholesterol&quot;</td>
<td><a href="http://rdf.rhea-db.org/33875">http://rdf.rhea-db.org/33875</a></td>
<td>&quot;cholesteryl (9Z-octadecenoate) + H2O = (9Z)-octadecenoate + cholesterol + H(+)}&quot;</td>
</tr>
<tr>
<td><a href="http://purl.obolibrary.org/obo/CHEBI_16113">http://purl.obolibrary.org/obo/CHEBI_16113</a></td>
<td>&quot;cholesterol&quot;</td>
<td><a href="http://rdf.rhea-db.org/34087">http://rdf.rhea-db.org/34087</a></td>
<td>&quot;a 1,2-diacyl-sn-glycero-3-phosphocholine + cholesterol = a 1-acyl-sn-glycero-3-phosphocholine + a cholesterol ester&quot;</td>
</tr>
</tbody>
</table>
Enzyme search

- Enzyme nomenclature (EC numbers)
- UniProt enzymes
Q5: Distribution of Rhea reactions based on the first level of IUBMB enzyme classification

```sparql
#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.-.-.-) }
    }
}
ORDER BY ?ecClass
```

<table>
<thead>
<tr>
<th>ecClass</th>
<th>ecClassName</th>
<th>reactionCount</th>
</tr>
</thead>
<tbody>
<tr>
<td><a href="http://purl.uniprot.org/enzyme/1.-.-.-">http://purl.uniprot.org/enzyme/1.-.-.-</a></td>
<td>Oxidoreductases</td>
<td>2445</td>
</tr>
<tr>
<td><a href="http://purl.uniprot.org/enzyme/2.-.-.-">http://purl.uniprot.org/enzyme/2.-.-.-</a></td>
<td>Transferases</td>
<td>2218</td>
</tr>
<tr>
<td><a href="http://purl.uniprot.org/enzyme/3.-.-.-">http://purl.uniprot.org/enzyme/3.-.-.-</a></td>
<td>Hydrolases</td>
<td>909</td>
</tr>
<tr>
<td><a href="http://purl.uniprot.org/enzyme/4.-.-.-">http://purl.uniprot.org/enzyme/4.-.-.-</a></td>
<td>Lyases</td>
<td>760</td>
</tr>
<tr>
<td><a href="http://purl.uniprot.org/enzyme/5.-.-.-">http://purl.uniprot.org/enzyme/5.-.-.-</a></td>
<td>Isomerases</td>
<td>322</td>
</tr>
<tr>
<td><a href="http://purl.uniprot.org/enzyme/6.-.-.-">http://purl.uniprot.org/enzyme/6.-.-.-</a></td>
<td>Ligases</td>
<td>224</td>
</tr>
<tr>
<td><a href="http://purl.uniprot.org/enzyme/7.-.-.-">http://purl.uniprot.org/enzyme/7.-.-.-</a></td>
<td>Translocases</td>
<td>76</td>
</tr>
</tbody>
</table>
Query: retrieve human enzyme metabolizing cholesterol and their reactions
Query: retrieve human enzyme metabolizing cholesterol and their reactions

1. NCBI taxid
   UniProt & Rhea
   CHEBI & Rhea

2. proteins
   reactions
   Proteins & Rhea reactions
   NCBI taxid (taxon:9606)

3. reactions
   Proteins & Rhea reactions
   NCBI taxid (taxon:9606)

4. HUMAN UniProt proteins
   Rhea reactions with compound CHEBI:16113
Q6: Retrieve human enzyme metabolizing cholesterol and their reactions

```sparql
#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/>
?reaction rdfs:subClassOf rh:Reaction .
# cholesterol (CHEBI:16113)
?compound rh:chebi ch:CHEBI_16113 .
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 .
?a1 a up:Catalytic_Activity_Annotation .
?a1 up:catalyticActivity ?ca .
?ca up:catalyzedReaction ?reaction ,
}
```
Q6: Retrieve human enzyme metabolizing cholesterol and their reactions

<table>
<thead>
<tr>
<th>proteinId</th>
<th>proteinName</th>
<th>reactionEquation</th>
</tr>
</thead>
<tbody>
<tr>
<td>&quot;DHCR7_HUMAN&quot;</td>
<td>&quot;7-dehydrocholesterol reductase&quot;</td>
<td>cholesterol + NAD(P) = cholesta-5,7-dien-3beta-ol + H(+) + NADPH</td>
</tr>
<tr>
<td>&quot;ABCD1_HUMAN&quot;</td>
<td>&quot;ATP-binding cassette sub-family G member 1&quot;</td>
<td>ATP + cholesterol(in) + H2O = ADP + cholesterol(out) + H(+) + phosphate</td>
</tr>
<tr>
<td>&quot;ABCD5_HUMAN&quot;</td>
<td>&quot;ATP-binding cassette sub-family G member 5&quot;</td>
<td>ATP + cholesterol(in) + H2O = ADP + cholesterol(out) + H(+) + phosphate</td>
</tr>
<tr>
<td>&quot;ABCD8_HUMAN&quot;</td>
<td>&quot;ATP-binding cassette sub-family G member 8&quot;</td>
<td>ATP + cholesterol(in) + H2O = ADP + cholesterol(out) + H(+) + phosphate</td>
</tr>
<tr>
<td>&quot;CEL_HUMAN&quot;</td>
<td>&quot;Bile salt-activated lipase&quot;</td>
<td>cholesteryl (9Z-octadecenoate) + H2O = (9Z)-octadecenoate + cholesterol + H(+)</td>
</tr>
<tr>
<td>&quot;CP46A_HUMAN&quot;</td>
<td>&quot;Cholesterol 24-hydroxylase&quot;</td>
<td>cholesterol + O2 + reduced [NADPH--hemoprotein reductase] = (24S)-hydroxycholesterol + H(+) +</td>
</tr>
<tr>
<td>&quot;CH23H_HUMAN&quot;</td>
<td>&quot;Cholesterol 25-hydroxylase&quot;</td>
<td>AH2 + cholesterol + O2 = 25-hydroxycholesterol + A + H2O</td>
</tr>
<tr>
<td>&quot;CH23H_HUMAN&quot;</td>
<td>&quot;Cholesterol 25-hydroxylase&quot;</td>
<td>cholesterol + H(+) + NADPH + O2 = 25-hydroxycholesterol + H2O + NADP(+)</td>
</tr>
<tr>
<td>&quot;CP11A_HUMAN&quot;</td>
<td>&quot;Cholesterol side-chain cleavage enzyme, mitochondrial&quot;</td>
<td>cholesterol + 2 H(+) + O2 + 2 reduced [adenodoxin] = 22R-hydroxycholesterol + H2O + 2 oxidized [adenodoxin]</td>
</tr>
<tr>
<td>&quot;CP11A_HUMAN&quot;</td>
<td>&quot;Cholesterol side-chain cleavage enzyme, mitochondrial&quot;</td>
<td>cholesterol + 6 H(+) + 3 O2 + 6 reduced [adenodoxin] = 4-methylpentanal + 4 H2O + 6 oxidized [adenodoxin]</td>
</tr>
</tbody>
</table>
# SPARQL query forms

<table>
<thead>
<tr>
<th>Query form</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>SELECT</td>
<td>Returns all, or a subset of, the variables bound in a query pattern match.</td>
</tr>
<tr>
<td>ASK</td>
<td>Returns a boolean indicating whether a query pattern matches or not.</td>
</tr>
<tr>
<td>CONSTRUCT</td>
<td>Returns an RDF graph constructed by substituting variables in a set of triple templates.</td>
</tr>
<tr>
<td>DESCRIBE</td>
<td>Returns an RDF graph that describes the resources found.</td>
</tr>
</tbody>
</table>

*(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)

**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?

```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 .
    ?compound rh:chebi ch:CHEBI_16113 .
  }# Human proteins (taxid=9606)
  ?protein up:organism taxon:9606 .
  # Rhea catalyzed reactions
  ?a1 a up:Catalytic_Activity_Annotation .
  ?a1 up:catalyticActivity ?ca .
  ?ca up:alyzedReaction ?reaction .
  # UniProt cellular components
  ?a2 a up:Subcellular_Location_Annotation .
  ?a2 up:locatedIn ?lIn .
  ?lIn up:cellularComponent ?upLocation .
  ?upLocation skos:prefLabel ?upLocationLabel .
  ?upLocation skos:exactMatch ?goId .
  ?a2 up:locatedIn ?lIn .
  ?lIn up:cellularComponent ?upLocation .
  ?upLocation skos:prefLabel ?upLocationLabel .
  ?upLocation skos:exactMatch ?gold .
}
```
Q8: Where are the human enzymes metabolizing cholesterol located in the cell?

Results

<table>
<thead>
<tr>
<th>protein</th>
<th>proteinId</th>
<th>reaction</th>
<th>upLocation</th>
<th>upLocationLabel</th>
<th>gold</th>
</tr>
</thead>
</table>

[...]
Query: where are the human genes encoding enzymes metabolizing cholesterol expressed?

1. NCBI taxid
2. Proteins & Rhea reactions
3. compound (CHEBI:16113)
4. Ensembl gene
5. Uberon anatomic entities
6. HUMAN UniProt proteins
   - Rhea reactions with compound CHEBI:16113
   - Human Ensembl genes
   - Uberon Anatomic entities

NCBI taxid
UniProt & Rhea
CHEBI & Rhea
UniProt & Ensembl
Ensembl & UBERON

bioSODA

HUMAN UniProt proteins
Rhea reactions with compound CHEBI:16113
Human Ensembl genes
Uberon Anatomic entities
Q9: Where are the human genes encoding enzymes metabolizing cholesterol expressed?

WHERE {
  SERVICE <https://sparql.rhea-db.org/sparql> {
    ?reaction rdfs:subClassOf rh:Reaction .
    ?compound rh:chebi ch:CHEBI_16113 .
  }
  ?protein up:organism taxon:9606 .
  ?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> {
  }
}
Q9: Where are the human genes encoding enzymes metabolizing cholesterol expressed?

### Results

<table>
<thead>
<tr>
<th>protein</th>
<th>ensemblGene</th>
<th>reaction</th>
<th>anatomicEntityLabel</th>
<th>anatomicEntity</th>
</tr>
</thead>
<tbody>
<tr>
<td>uniprot:P06684</td>
<td><a href="http://rdf.ebi.ac.uk/resource/ensembl/ENSG00000160686">http://rdf.ebi.ac.uk/resource/ensembl/ENSG00000160686</a></td>
<td>rh:46140</td>
<td><em>left testis</em></td>
<td>obo:UBERON_0004533</td>
</tr>
<tr>
<td>uniprot:P06684</td>
<td><a href="http://rdf.ebi.ac.uk/resource/ensembl/ENSG00000160686">http://rdf.ebi.ac.uk/resource/ensembl/ENSG00000160686</a></td>
<td>rh:46140</td>
<td><em>pituitary gland</em></td>
<td>obo:UBERON_0000007</td>
</tr>
<tr>
<td>uniprot:P06684</td>
<td><a href="http://rdf.ebi.ac.uk/resource/ensembl/ENSG00000160686">http://rdf.ebi.ac.uk/resource/ensembl/ENSG00000160686</a></td>
<td>rh:46140</td>
<td><em>anterior cingulate cortex</em></td>
<td>obo:UBERON_0009835</td>
</tr>
<tr>
<td>uniprot:P06684</td>
<td><a href="http://rdf.ebi.ac.uk/resource/ensembl/ENSG00000160686">http://rdf.ebi.ac.uk/resource/ensembl/ENSG00000160686</a></td>
<td>rh:46140</td>
<td><em>ectocervix</em></td>
<td>obo:UBERON_0012249</td>
</tr>
<tr>
<td>uniprot:P06684</td>
<td><a href="http://rdf.ebi.ac.uk/resource/ensembl/ENSG00000160686">http://rdf.ebi.ac.uk/resource/ensembl/ENSG00000160686</a></td>
<td>rh:46140</td>
<td><em>putamen</em></td>
<td>obo:UBERON_0001874</td>
</tr>
<tr>
<td>uniprot:P06684</td>
<td><a href="http://rdf.ebi.ac.uk/resource/ensembl/ENSG00000160686">http://rdf.ebi.ac.uk/resource/ensembl/ENSG00000160686</a></td>
<td>rh:46140</td>
<td><em>right lung</em></td>
<td>obo:UBERON_0002167</td>
</tr>
<tr>
<td>uniprot:P06684</td>
<td><a href="http://rdf.ebi.ac.uk/resource/ensembl/ENSG00000160686">http://rdf.ebi.ac.uk/resource/ensembl/ENSG00000160686</a></td>
<td>rh:46140</td>
<td><em>esophagogastric junction muscularis propria</em></td>
<td>obo:UBERON_0035841</td>
</tr>
<tr>
<td>uniprot:P06684</td>
<td><a href="http://rdf.ebi.ac.uk/resource/ensembl/ENSG00000160686">http://rdf.ebi.ac.uk/resource/ensembl/ENSG00000160686</a></td>
<td>rh:46140</td>
<td><em>right ovary</em></td>
<td>obo:UBERON_0002118</td>
</tr>
<tr>
<td>uniprot:P06684</td>
<td><a href="http://rdf.ebi.ac.uk/resource/ensembl/ENSG00000160686">http://rdf.ebi.ac.uk/resource/ensembl/ENSG00000160686</a></td>
<td>rh:46140</td>
<td><em>aorta</em></td>
<td>obo:UBERON_0000947</td>
</tr>
<tr>
<td>uniprot:P06684</td>
<td><a href="http://rdf.ebi.ac.uk/resource/ensembl/ENSG00000160686">http://rdf.ebi.ac.uk/resource/ensembl/ENSG00000160686</a></td>
<td>rh:46140</td>
<td><em>minor salivary gland</em></td>
<td>obo:UBERON_0001830</td>
</tr>
<tr>
<td>uniprot:P06684</td>
<td><a href="http://rdf.ebi.ac.uk/resource/ensembl/ENSG00000160686">http://rdf.ebi.ac.uk/resource/ensembl/ENSG00000160686</a></td>
<td>rh:46140</td>
<td><em>subcutaneous adipose tissue</em></td>
<td>obo:UBERON_0002190</td>
</tr>
<tr>
<td>uniprot:P06684</td>
<td><a href="http://rdf.ebi.ac.uk/resource/ensembl/ENSG00000160686">http://rdf.ebi.ac.uk/resource/ensembl/ENSG00000160686</a></td>
<td>rh:46140</td>
<td><em>blood</em></td>
<td>obo:UBERON_0000178</td>
</tr>
</tbody>
</table>
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](https://sparql.rhea-db.org/sparql)

```sparql
PREFIX rh:<http://rdf.rhea-db.org/>
PREFIX up:<http://purl.uniprot.org/core/>
PREFIX taxon:<http://purl.uniprot.org/taxonomy/>

WHERE {
  SERVICE <http://sparql.uniprot.org/sparql> {
    # Filter by NCBI taxid (H. pylori == 85962)
    ?protein up:organism taxon:85962 .
    ?protein up:mnemonic ?proteinId .
    # Rhea reactions catalyzed by UniProt proteins
    ?a up:catalyticActivity ?ca .
    ?ca up:catalyzedReaction ?reaction .

    ?reaction rh:equation ?reactionEquation
    ?reaction rh:side ?reactionSide
    ?reactionSide rh:contains ?reactionParticipant
    ?reactionParticipant rh:compound ?compound
    OPTIONAL {?compound rh:chebi ?chebi }.
  }
}
```
Q10: *H. pylori* metabolic network (UniProtKB/Swiss-Prot and Rhea-ChEBI)

<table>
<thead>
<tr>
<th>protein</th>
<th>proteinId</th>
<th>reaction</th>
<th>reactionSide</th>
<th>compound</th>
<th>chebi</th>
<th>reactionEquation</th>
</tr>
</thead>
</table>

[...]  

→ To compare with published GSMNs (MetaNetX, Marco)
Q11: Explore *H. pylori* tryptophan biosynthesis pathway (GO:0000162)

### SPARQL Query

```
#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/oboGO-> 

SELECT distinct ?upProteinId ?goId ?rhReaction ?upPathway 
WHERE { 
?upProtein up:reviewed true . 
?upProtein up:mnemonic ?upProteinId . 
# HELPY proteins (taxid=85962) 
?upProtein up:organism taxon:85962 . 
?upProtein up:classificationWith ?goId . 
VALUES ?goId {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 
```
Q12: *H. pylori* enzyme complexes for tryptophan biosynthesis pathway (MetaNetX: seed_Opt85962_1)
Q12: *H. pylori* enzyme complexes for tryptophan biosynthesis pathway (MetaNetX: seed_Opt85962_1)

<table>
<thead>
<tr>
<th>upPathway</th>
<th>upProteinId</th>
<th>rhReaction</th>
<th>mnxr</th>
<th>cplx_label</th>
<th>mnet</th>
</tr>
</thead>
<tbody>
<tr>
<td>Amino-acid biosynthesis; L-tryptophan biosynthesis; L-tryptophan from chorismate: step 2/5</td>
<td>TRPD_HELRY</td>
<td><a href="http://rdf.rhea-db.org/11768">http://rdf.rhea-db.org/11768</a></td>
<td><a href="https://rdf.metanetx.org/reac/MNXR95842">https://rdf.metanetx.org/reac/MNXR95842</a></td>
<td>&quot;bact:TRPD_HELPRY&quot;</td>
<td></td>
</tr>
</tbody>
</table>
Programmatic access

**SPARQLWrapper**
- pandas
- matplotlib
- matplotlib_venn

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

```
<table>
<thead>
<tr>
<th>category</th>
<th>label</th>
<th>count</th>
<th>pct</th>
</tr>
</thead>
<tbody>
<tr>
<td>domain</td>
<td>archaea</td>
<td>857</td>
<td>8.6%</td>
</tr>
<tr>
<td>domain</td>
<td>bacteria</td>
<td>3334</td>
<td>33.6%</td>
</tr>
<tr>
<td>domain</td>
<td>eukaryota</td>
<td>4221</td>
<td>42.5%</td>
</tr>
<tr>
<td>domain</td>
<td>viruses</td>
<td>187</td>
<td>1.9%</td>
</tr>
</tbody>
</table>
```

Taxonomic distribution of the Rhea reactions annotated in UniProtKB/Swiss-Prot
Archaea: 774; Bacteria: 3436; Eukaryota: 4080
(release: Rhea 108 / UniProt 2019_09)
2.1 Python Libraries

from SPARQLWrapper import SPARQLWrapper, JSON
from pandas.io.json import json_normalize

2.2 SPARQL endpoints

sparql_uniprot_url = 'https://sparql.uniprot.org/sparql/
sparsl_rhea_url = 'https://sparql.rhea-db.org/sparql/

2.3 Python functions

```python
def sparql2pandas(sparql_query, sparql_service_url):
    ... # 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
sparql q1=""
#endpoint: https://sparql.rhea-db.org/sparql
PRELIF rdf:<http://rdf.rhea-db.org/>
SELECT ?reaction ?reactionEquation
WHERE {
    ?reaction rdf:subClassOf rhd:Reaction .
    VALUES ?status (rhd:Approved rhd:Preliminary) .
}
ORDER BY ?reaction
print(sparql q1)
```

4.1.2 Compute Venn diagram using matplotlib and matplotlib_venn python libraries

```python
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']['rhd:Reaction'].drop_duplicates())
set_B = set(df_up_rh[df_up_rh.domainName == 'Bacteria']['rhd:Reaction'].drop_duplicates())
set_X = set(df_up_rh[df_up_rh.domainName == 'Eukaryota']['rhd:Reaction'].drop_duplicates())

nb_A = len(set_A)
nb_B = len(set_B)
nb_X = len(set_X)
nb_all = len(set_A | set_B | set_X)

title = 'Taxonomic distribution of the Rhea reactions annotated in UniProtKB/Swiss-Prot\n'Rhea reactions: in Archaea=' + str(nb_A) + '; in Bacteria=' + str(nb_B) + '; in Eukaryota=' + str(nb_X) + '\n
# Plot Venn diagram
venn3([set_A, set_B, set_X], set_labels = ('Bacteria', 'Eukaryota', 'Archaea'))
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

01  What is Rhea?
02  Data model
03  SPARQLing Rhea
04  Summary
Summary

How to query metabolic data across multiple data source:

<table>
<thead>
<tr>
<th>Source</th>
<th>Description</th>
<th>Web site</th>
<th>SPAQRL endpoint</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rhea</td>
<td>chemical reactions</td>
<td><a href="https://www.rhea-db.org/">https://www.rhea-db.org/</a></td>
<td><a href="https://sparql.uniprot.org/sparql/">https://sparql.uniprot.org/sparql/</a></td>
</tr>
</tbody>
</table>

- 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

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
Thank you for your attention

anne.morgat@sib.swiss