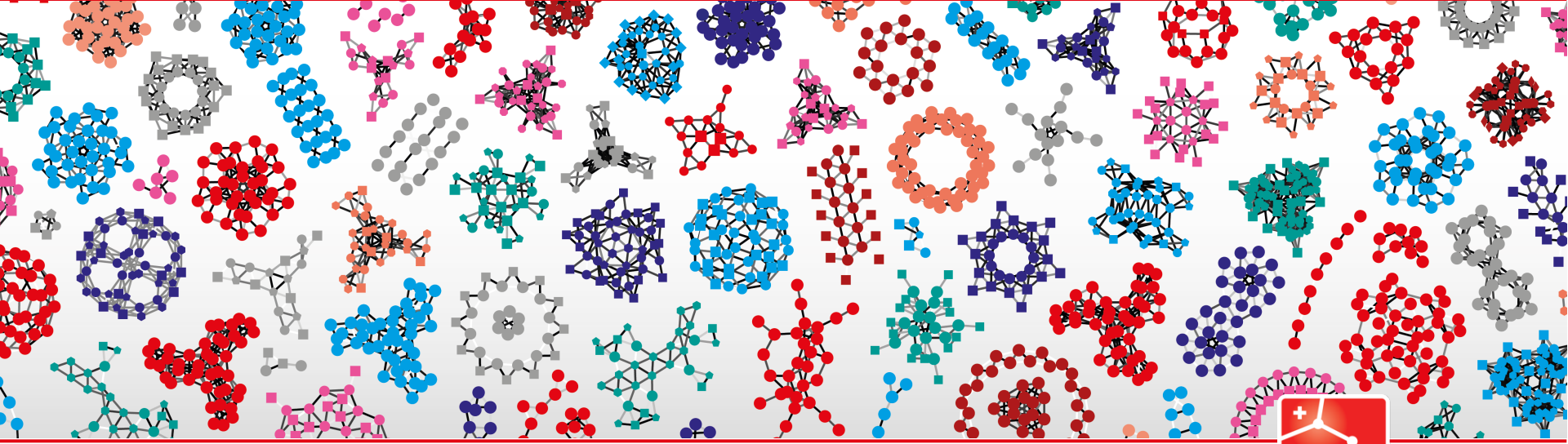


Swiss Institute of Bioinformatics

How to mine enzyme data ?

Practicals (and corrections)

Anne Morgat, Swiss-Prot group, SIB Swiss Institute of Bioinformatics

elixir SWITZERLAND | swissprot | Rhea

Lausanne, March 3rd 2020

www.sib.swiss

Q1: Retrieve human enzymes (annotated with EC numbers)

- How many human enzymes in UniProt ?
- How many expert-curated human enzymes?

Answer: Retrieve human enzymes (annotated with EC numbers)

- How many human enzymes in UniProt ? 14672
  ec:* AND organism:"Homo sapiens (Human) [9606]"
- How many expert-curated human enzymes (UniProtKB/Swiss-Prot)? 4345
  ec:* AND reviewed:yes AND organism:"Homo sapiens (Human) [9606]"

If you have questions related to the HUMAN proteome, please read https://www.uniprot.org/help/human_proteome

Q2: Retrieve human enzymes with Catalytic activity annotations

- How many human enzymes in UniProtKB/Swiss-Prot have Catalytic activity annotations ?

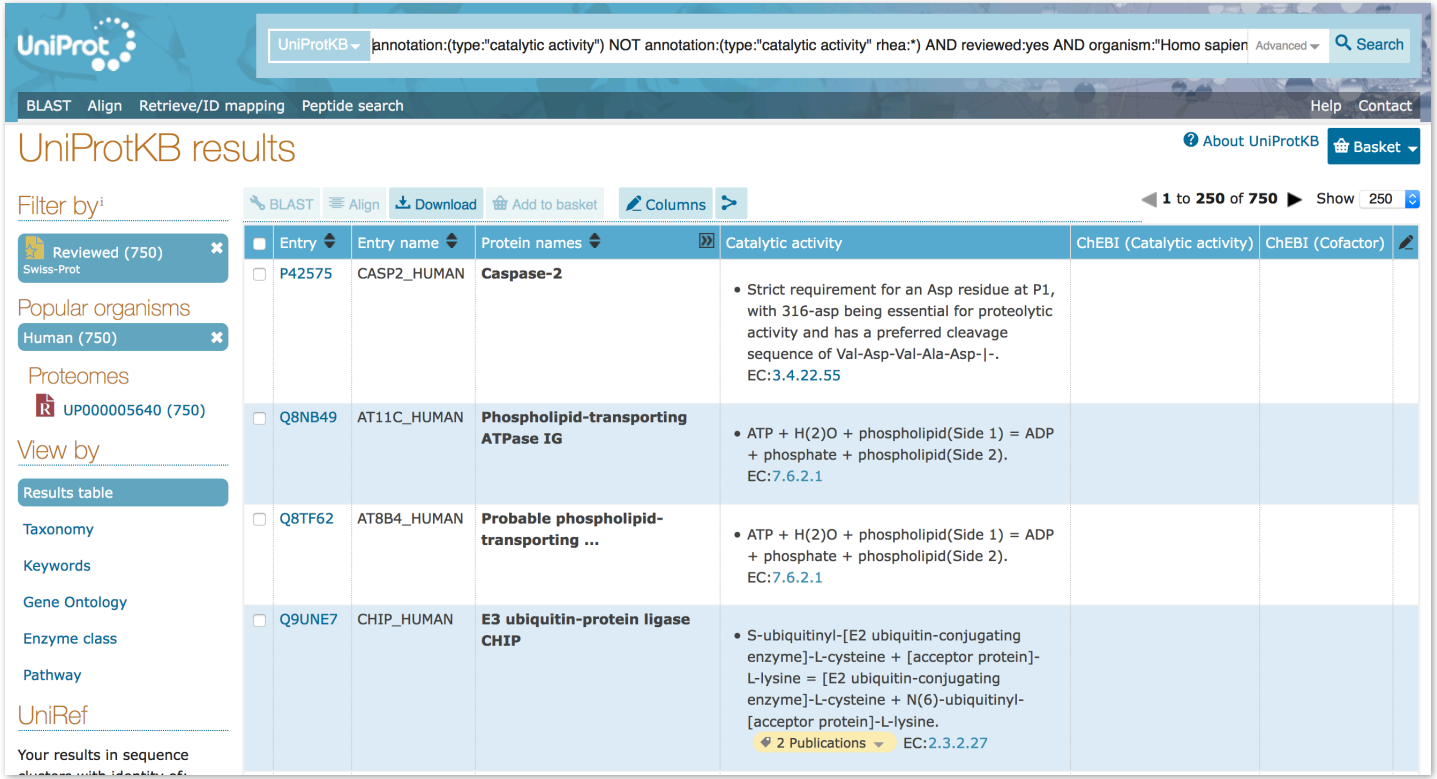
- How many human enzymes in UniProtKB/Swiss-Prot have Catalytic activity annotations based on Rhea reactions?

- How do you explain the difference?

Answer: Retrieve human enzymes with Catalytic activity annotations (1)

- How many human enzymes in UniProtKB/Swiss-Prot have Catalytic activity annotations ? 3,617
  Query: annotation:(type:"catalytic activity") AND reviewed:yes AND organism:"Homo sapiens (Human) [9606]"
- How many human enzymes in UniProtKB/Swiss-Prot have Catalytic activity annotations based on Rhea reactions? 2,867
  Query: annotation:(type:"catalytic activity" rhea:*) AND reviewed:yes AND organism:"Homo sapiens (Human) [9606]"
- How do you explain the difference?
  There are 750 HUMAN UniProtKB/Swiss-Prot that have a Catalytic activity annotation without Rhea reactions
  Query: annotation:(type:"catalytic activity") NOT annotation:(type:"catalytic activity" rhea:*) AND organism:"Homo sapiens (Human) [9606]" AND reviewed:yes

Answer: Retrieve human enzymes with Catalytic activity annotations (2)



Q3: Search UniProt by chemical name: ‘cholesterol’

Simple search

- How many UniProt entries?
- How many Swiss-Prot entries?
- How many TrEMBL entries?

Advanced search (Catalytic Activity)

- How many Swiss-Prot entries?
- How many Swiss-Prot entries with any experimental evidence?
- How many HUMAN entries with any experimental evidence?

Answer: Search UniProt by chemical name: ‘cholesterol’

Simple search

- How many UniProt entries? 156,727
- How many Swiss-Prot entries? 2354
- How many TrEMBL entries? 154373

Advanced search (Catalytic Activity)

- How many Swiss-Prot entries? 131
- How many Swiss-Prot entries with any experimental evidence? 47
- How many HUMAN entries with any experimental evidence? 20

Q4: Search UniProtKB/Swiss-Prot by chemical name: ‘NAD+’

Using UniProt Advanced search for small molecule

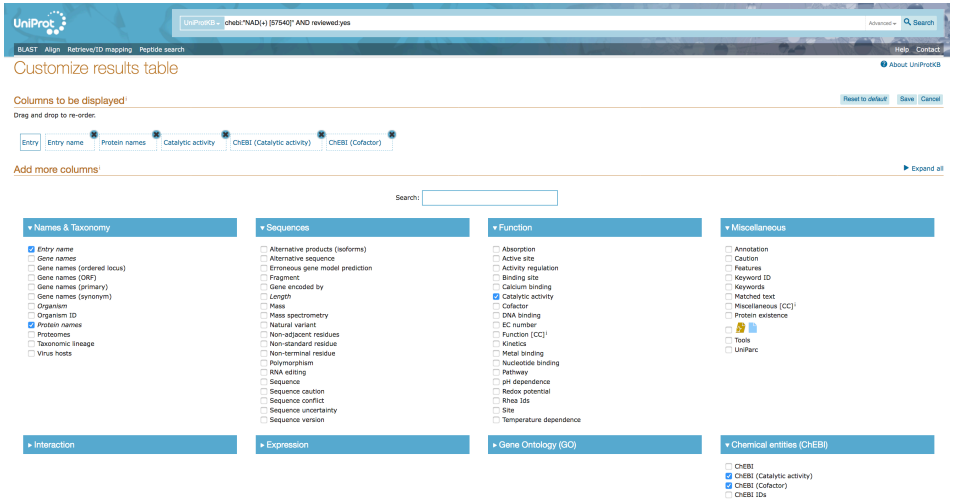
- What is the ChEBI Id of NAD+?
- How many Swiss-Prot entries are annotated with NAD+?
- What is the role of NAD+?
  Tips: Customize Result table to display:
  Entry name, Protein names, Catalytic activity, ChEBI (Catalytic activity), ChEBI (Cofactor)
- Can NAD+ be both a cofactor and a reaction participant in the same enzyme?

Answer: Search UniProt by chemical name: ‘NAD+’ (1)

Advanced search (small molecule)

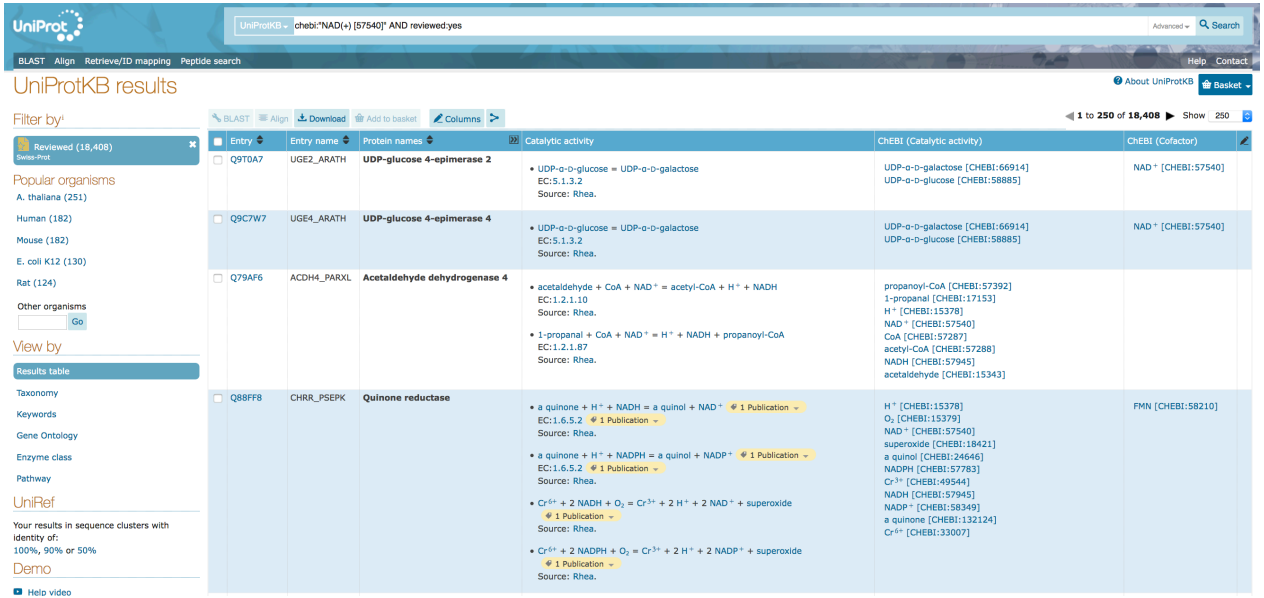
- What is the ChEBI ID? CHEBI:57540
- How many Swiss-Prot entries? 18,408
  chebi:"NAD(+) [57540]" AND reviewed:yes

- Customize Result table to display:
  Entry name
  Protein names
  Catalytic activity
  ChEBI (Catalytic activity)
  ChEBI (Cofactor)



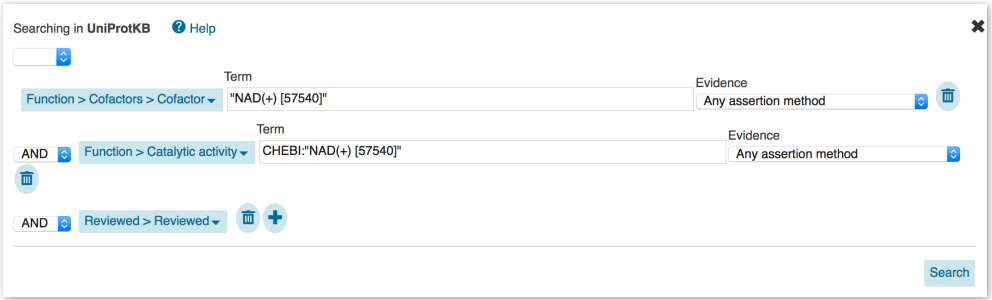
Answer: Search UniProt by chemical name: ‘NAD+’ (2)

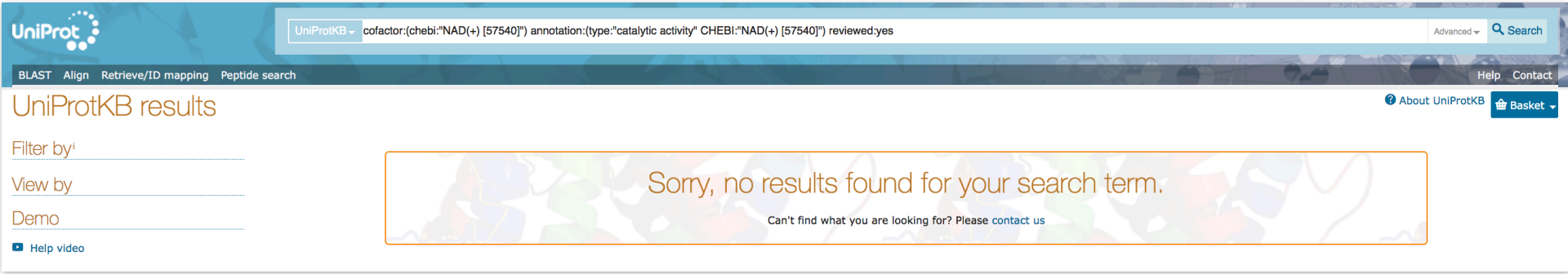
- What do you observe? NAD(+) [CHEBI:57540] is either a reaction participant or a cofactor



Answer: Search UniProt by chemical name: ‘NAD+’ (3)

- Can NAD+ be both a cofactor <u>and</u> a reaction participant in the same enzyme? no





Q5: Search UniProtKB/Swiss-Prot entries for RHEA:19253

Advanced search (Catalytic activity)

- How many UniProtKB/Swiss-Prot entries?
- What are the corresponding reactions in KEGG, MetaCyc and Reactome?
- What are parent/children reactions of RHEA:19253?

Answer: Search UniProtKB/Swiss-Prot entries for RHEA:19253

Advanced search (Catalytic activity)

- How many UniProtKB/Swiss-Prot entries? 34
- What are the corresponding reactions in KEGG, MetaCyc and Reactome?
  Go to https://www.rhea-db.org/reaction?id=19253
- What are parent/children reactions of RHEA:19253?
  Parent reaction:
  - RHEA:45300 an N-(acyl)-sphingosylphosphocholine + H2O = H(+) + N-acyl-sphingoid base + phosphocholine

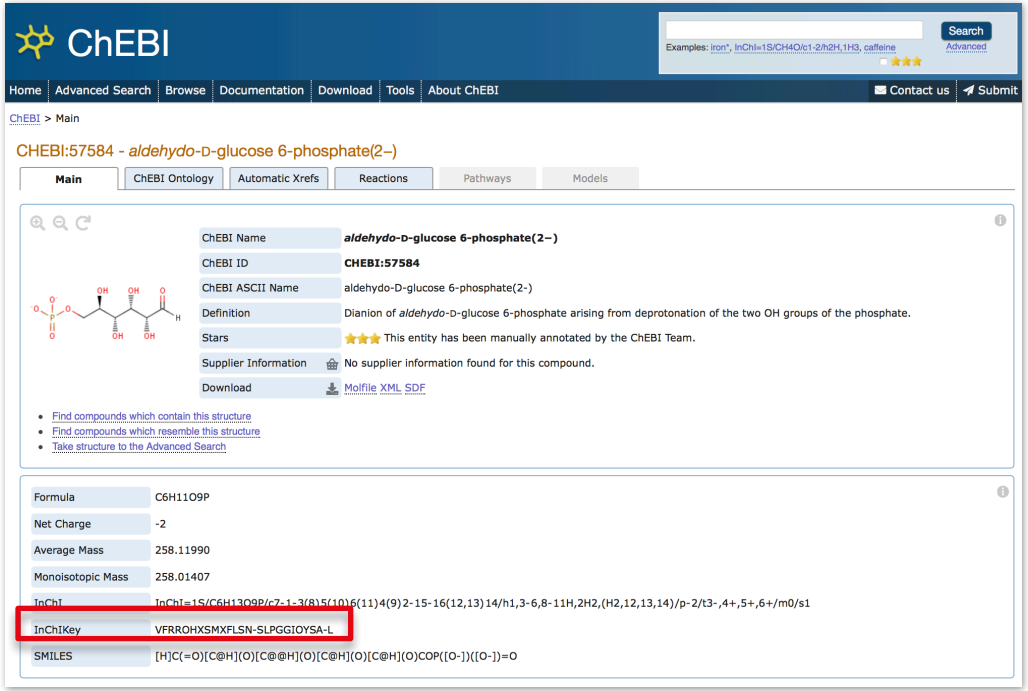
  Children reactions
  - RHEA:45324H2O + N-(tetracosanoyl)-sphing-4-enine-1-phosphocholine = H(+) + N-(tetracosanoyl)-sphing-4-enine + phosphocholine
  - RHEA:54284H2O + N-(octadecanoyl)-sphing-4-enine-1-phosphocholine = H(+) + N-octadecanoylsphing-4-enine + phosphocholine

Q6: Search by InChIKey

- Search ChEBI:57584 in ChEBI resource and copy its InChiKey
- How many UniProt entries are annotated with the full InChIKey?
- Remove the last two part, and see how many UniProt entries are found?

Answer: Search by InChIKey (1)

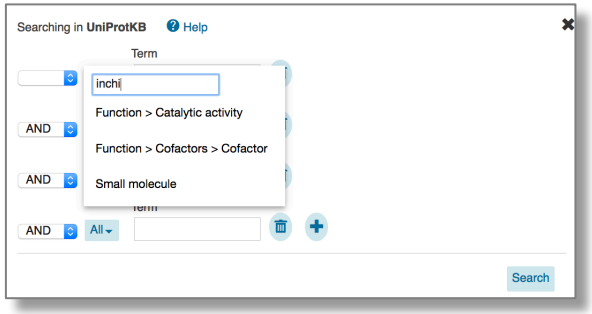
- Search ChEBI:57584 in ChEBI resource and copy its InChiKey
  https://www.ebi.ac.uk/chebi/searchId.do?chebiId=CHEBI:57584
  VFRROHXSMXFLSN-SLPGGIOYSA-L

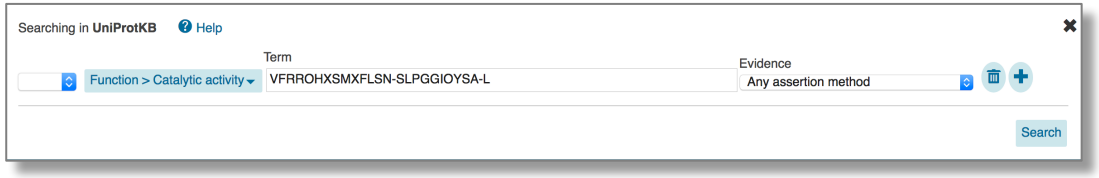


Answer: Search by InChIKey (2)

- How many UniProt entries are annotated with the full InChIKey? 38,647

Advanced search
Term “inchi” – Select Function > Catalytic activity





Query: annotation:(type:"catalytic activity" inchikey:VFRROHXSMXFLSN-SLPGGIOYSA-L)
640 reviewed (UniProtKB/Swiss-Prot)
38,007 unreviewed (UniProtKB/TrEMBL)

Answer: Search by InChIKey (3)

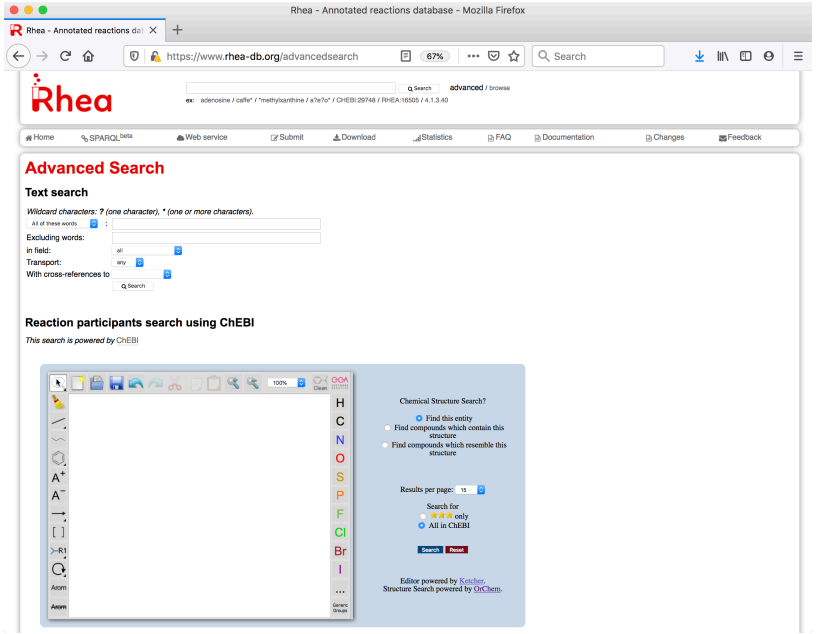
- Remove the last two part, and see how many UniProt entries are found? 40,458
  Query: annotation:(type:"catalytic activity" inchikey:VFRROHXSMXFLSN)
  754 reviewed (UniProtKB/Swiss-Prot)
  39,704 unreviewed (UniProtKB/TrEMBL)

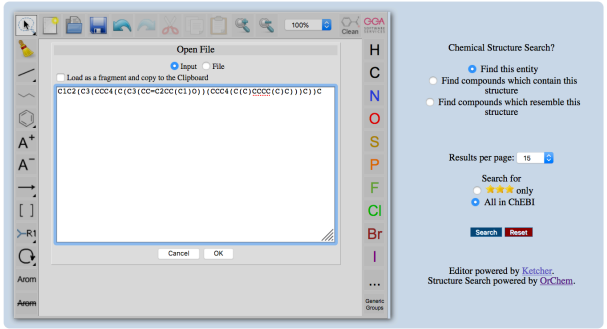
Q6: Retrieve cholesterol and derivatives using structural search

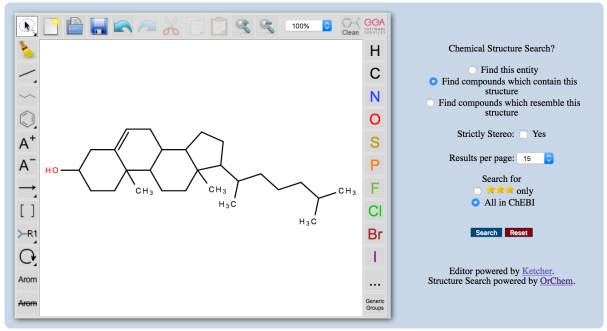
- Search cholesterol in ChEBI and copy its SMILES
- Perform a substructural search in the Rhea web site
- How many cholesterol or cholesterol derivatives are use in Rhea?
- Retrieve the Rhea reaction(s) involving protodioscin
- Retrieve their catalyst(s) in UniProtKB/Swiss-Prot

Answer: Retrieve cholesterol and derivatives using structural search(1)

- Search cholesterol in ChEBI resource and copy its SMILES
  C1C2(C3(CCC4(C(C3(CC=C2CC(C1)O))(CCC4(C(C)CCCC(C)C)))C))C
- Rhea web site - Advanced search

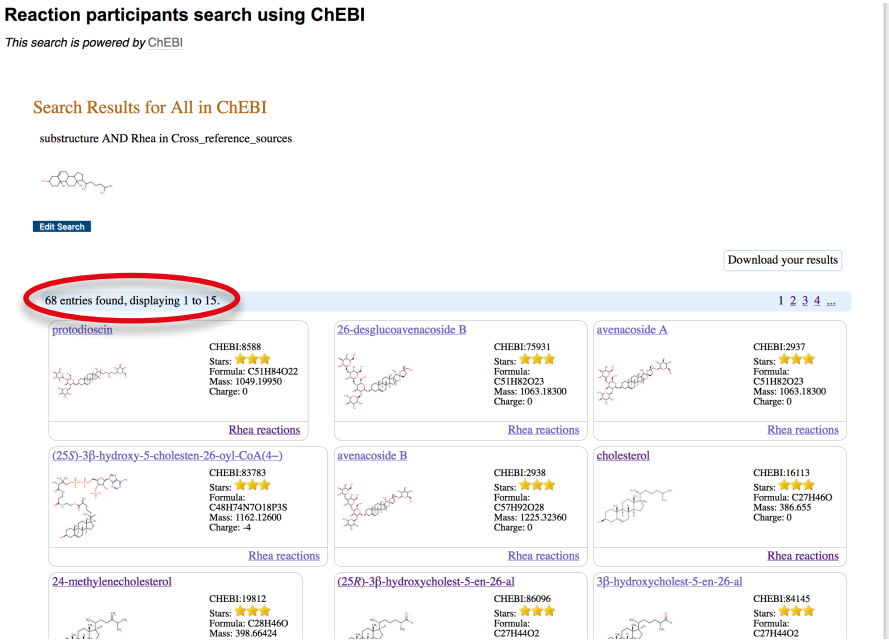


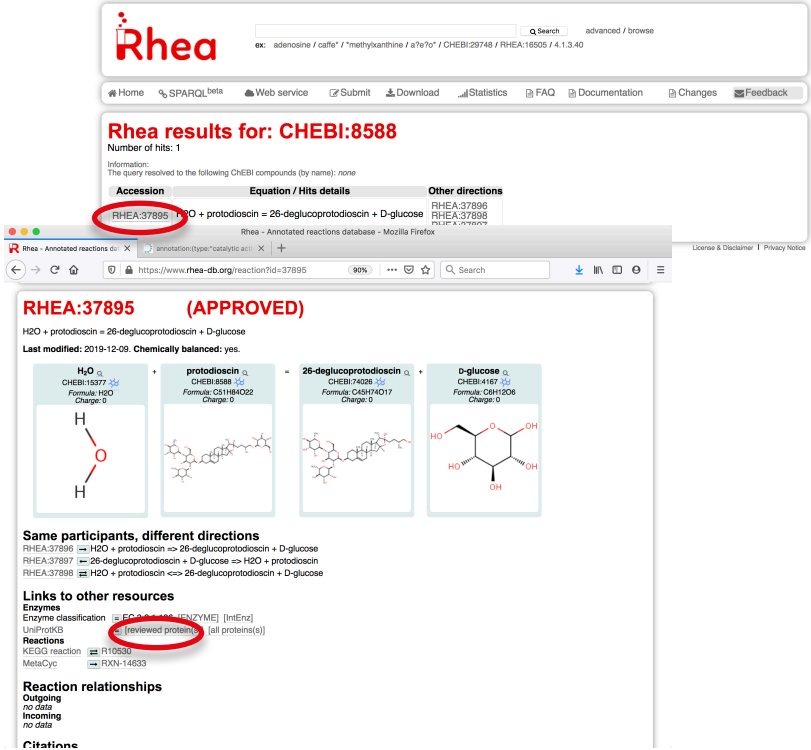




Answer: Retrieve cholesterol and derivatives using structural search(2)

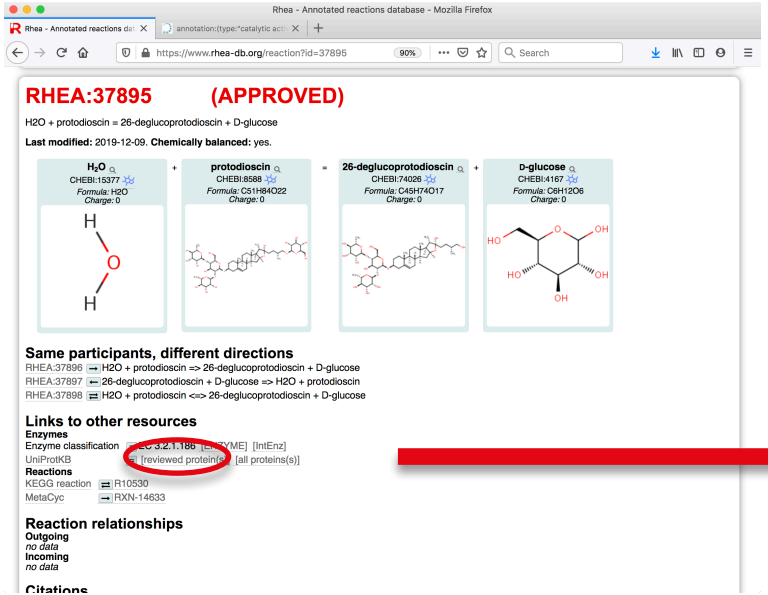
- How many cholesterol or cholesterol derivatives are use in Rhea? 68
- Retrieve the Rhea reaction(s) involving protodioscin

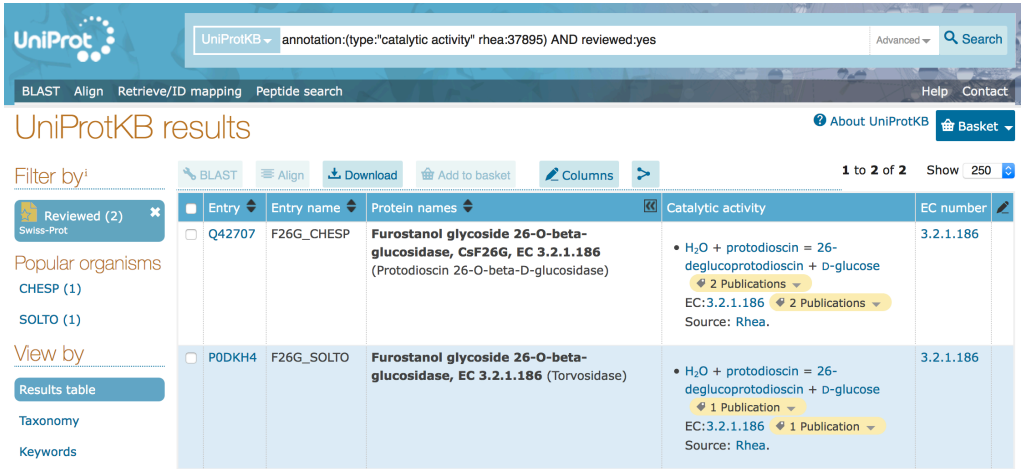




Answer: Retrieve cholesterol and derivatives using structural search(3)

- Retrieve their catalyst(s) in UniProtKB/Swiss-Prot





A case study (1)

« Biologists analyzing a metabolomics dataset from a cardiovascular disease cohort finds alterations in metabolites whose masses are consistent with cholesterol and related molecules. They also have proteomics data that show significant differences in some proteins. »

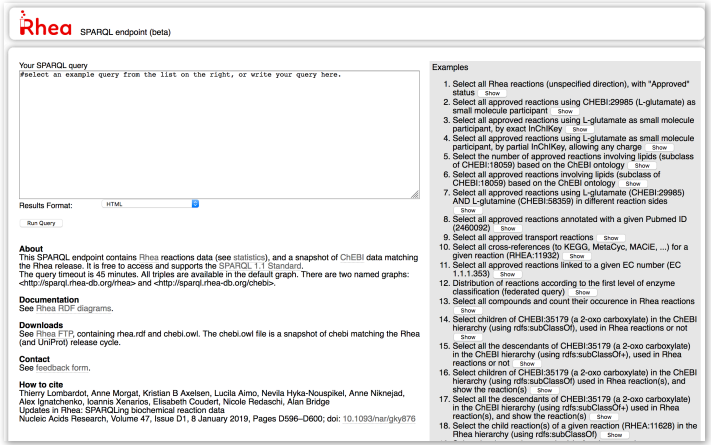
Questions:

- What are the enzymes involved in human cholesterol metabolism?
- Are the enzymes known to be linked to cardiovascular diseases (CVD) or other diseases?

SPARQL query: Retrieve the Rhea biochemical reactions that involve cholesterol or cholesterol derivatives (using sachem:substructureSearch)

Cholesterol SMILES = C1C2(C3(CCC4(C(C3(CC=C2CC(C1)O))(CCC4(C(C)CCCC(C)C)))C))C

```
# endpoint: https://sparql.rhea-db.org/sparql
PREFIX ch:http://purl.obolibrary.org/obo/
PREFIX taxon:http://purl.uniprot.org/taxonomy/
PREFIX keywords:http://purl.uniprot.org/keywords/
SELECT distinct ?CHEBI
                ?CHEBI_UNIPROT_NAME
                ?RHEA
                ?REACTION_EQUATION
WHERE {
  # idsm:chebi service
  SERVICE idsm:chebi {
    ?CHEBI sachem:substructureSearch [ sachem:query
                'C1C2(C3(CCC4(C(C3(CC=C2CC(C1)O))(CCC4(C(C)CCCC(C)C)))C))C' ] .
  }
  ?RHEA_REACTION rdfs:subClassOf rh:Reaction .
  ?RHEA_REACTION rh:status rh:Approved .
  ?RHEA_REACTION rh:side / rh:contains / rh:compound / rh:chebi ?CHEBI .
  ?CHEBI up:name ?CHEBI_UNIPROT_NAME .
}
```



Copy/paste the query in https://sparql.rhea-db.org/sparql

A case study (2)

Query: Retrieve the number of UniProtKB/Swiss-Prot human enzymes that metabolize cholesterol or cholesterol derivatives (using sachem:substructureSearch)
See query example 22

Query: Retrieve the list of UniProtKB/Swiss-Prot human enzymes that metabolize cholesterol or cholesterol derivatives (using sachem:substructureSearch)
See query example 23

Query: Retrieve the number of UniProtKB/Swiss-Prot human enzymes that metabolize cholesterol or cholesterol derivatives and that are involved in diseases (federated query using Rhea, IDSM/Sachem and UniProt)
See query example 24

Query: Retrieve the list of diseases involving human enzymes that metabolize cholesterol or cholesterol derivatives and the number of proteins involved (federated query using Rhea, IDSM/Sachem and UniProt)
See query example 25