* Documentation about ChEMBL end-point RDF

3September

In case of very large output how can we reach to the object URI for further analysis of object source where object source are not common type?????

Skos:label has many information so if We want to print a single value then ????

I did not find the mechanism of action of molecules.

Mol 1886521-192 (tautomer+ enantiomer)

Tautomer are generally consider as a same compound.

A positive and negative charge possible on adjacent atom????

Mol 192-168949 diff str ..fine!!! but why not alt. configuration????

4September

Some molecules name have wrong labelling.

In SQL Server CHEMBL33780 IC50 40&44 . Page containing this information has 40 and many others value but not 44.

Where is the molregne????

Target molecules have wrong labelling so how can I choose a particular target????

Some of properties did not get????

Some of questions are unclear mentioned on CHEMBL webpage????

5-6Sept: weekend

7September

LIMIT is not working.

Stuck with INTERSECT problem using SPARQL????

8September

hasTarget URI used at two places so wondering about ambiguity????

Target Type is just the string, not URI so unable to select type.

Accession no. has mentioned as uniprot reference.

Pubmed-id has just the link not the labelling.

NOTE: Some of the value are null like std-relation, std-type, std-unit so best way to do query is put optional if needed.

dct prefix shown in schema is equivalent to dcterms used in interface.

- Competency questions:

- Knowing that the identifier of sildenafil is CHEMBL192 (<https://www.ebi.ac.uk/chembl/compound/inspect/CHEMBL192>):

- Can we find the resource to the identifier in the triple store?

<http://rdf.ebi.ac.uk/resource/chembl/molecule/CHEMBL192>

- If I only know that the molecule is called “sildenafil”, how do I find the resource/URI?

SELECT ?molecule

WHERE {

 ?molecule rdfs:label "SILDENAFIL" .

}

How to find the molecules or sources for a combination of molecular farmula “C22H30N6O4S”?

SELECT ?molecule

WHERE {

 ?molecule rdfs:label ?o.

 FILTER regex(?o, "C22H30N6O4S", "i")

}

comment: It takes longer time but comes with solution.

- Can we find the molecular formula in the triple store?

SELECT ?molfar

WHERE {

 <http://rdf.ebi.ac.uk/resource/chembl/molecule/CHEMBL192#full\_molformula> ?p ?molfar.

}

- What are the trade names?

SELECT   ? tradename

WHERE {

 <http://rdf.ebi.ac.uk/resource/chembl/molecule/CHEMBL192> skos:altLabel ? tradename.

}

- What’s the ALogP?

SELECT  ? logpvalue

WHERE {

 <http://rdf.ebi.ac.uk/resource/chembl/molecule/CHEMBL192#alogp> ?p ? logpvalue.

}

- What’s the number of rotatable bounds?

SELECT  ? rotbon

WHERE {

 <http://rdf.ebi.ac.uk/resource/chembl/molecule/CHEMBL192#rtb> ?p ? rotbon.

}

- What’s the mechanism of action?

- Canonical SMILES? Is there other SMILES string available?

[<http://rdf.ebi.ac.uk/resource/chembl/molecule/CHEMBL192#canonical\_smiles>](http://www.ebi.ac.uk/rdf/services/chembl/describe?uri=http%3A%2F%2Frdf.ebi.ac.uk%2Fresource%2Fchembl%2Fmolecule%2FCHEMBL192%23canonical_smiles) ?p ?o.

- What are the activities known inside ChEMBL involving this compound?

SELECT ?act

WHERE {

 <http://rdf.ebi.ac.uk/resource/chembl/molecule/CHEMBL192> cco:hasActivity ?act.

}

furtheranlysisof activity using:

SELECT ?p ?o

WHERE {

 <http://rdf.ebi.ac.uk/resource/chembl/activity/CHEMBL\_ACT\_837350> ?p ?o .

}

**Retrieve all the bioactivity data for bacterial targets:**

PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>

PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>

PREFIX owl: <http://www.w3.org/2002/07/owl#>

PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>

PREFIX dc: <http://purl.org/dc/elements/1.1/>

PREFIX dcterms: <http://purl.org/dc/terms/>

PREFIX dbpedia2: <http://dbpedia.org/property/>

PREFIX dbpedia: <http://dbpedia.org/>

PREFIX foaf: <http://xmlns.com/foaf/0.1/>

PREFIX skos: <http://www.w3.org/2004/02/skos/core#>

PREFIX cco: <http://rdf.ebi.ac.uk/terms/chembl#>

SELECT ?molecule

WHERE {

?molecule rdfs:subClassOf cco:Substance.

?molecule cco:hasActivity ?act.

 ?act cco:hasAssay ?ass.

 ?ass cco:hasTarget ?tar.

 ?tar cco:organismName 'Bacteria'.

}

**Retrieve all the compounds which have an IC50 bioactivity value in nM:**

PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>

PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>

PREFIX owl: <http://www.w3.org/2002/07/owl#>

PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>

PREFIX dc: <http://purl.org/dc/elements/1.1/>

PREFIX dcterms: <http://purl.org/dc/terms/>

PREFIX dbpedia2: <http://dbpedia.org/property/>

PREFIX dbpedia: <http://dbpedia.org/>

PREFIX foaf: <http://xmlns.com/foaf/0.1/>

PREFIX skos: <http://www.w3.org/2004/02/skos/core#>

PREFIX cco: <http://rdf.ebi.ac.uk/terms/chembl#>

SELECT  ?chembl\_id  ?std\_value ?std\_unit  ?std\_type

WHERE {

?molecule cco:hasActivity ?Activity.

?molecule rdfs:subClassOf cco:Substance.

?Activity cco:standardValue ?std\_value .

?Activity cco:standardUnits ?std\_unit .

?Activity cco:standardType ?std\_type .

?molecule rdfs:label ?chembl\_id .

FILTER regex(?std\_unit , "nM").

FILTER regex(?std\_type, "IC50") .

}

**Retrieve compound activity details for a target:**

PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>

PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>

PREFIX owl: <http://www.w3.org/2002/07/owl#>

PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>

PREFIX dc: <http://purl.org/dc/elements/1.1/>

PREFIX dcterms: <http://purl.org/dc/terms/>

PREFIX dbpedia2: <http://dbpedia.org/property/>

PREFIX dbpedia: <http://dbpedia.org/>

PREFIX foaf: <http://xmlns.com/foaf/0.1/>

PREFIX skos: <http://www.w3.org/2004/02/skos/core#>

PREFIX cco: <http://rdf.ebi.ac.uk/terms/chembl#>

SELECT ?chembl\_id ?std\_value ?std\_unit ?std\_type ?std\_rel ?ass\_disc

WHERE {

 ?molecule cco:hasActivity ?act.

?molecule rdfs:subClassOf cco:Substance.

 ?molecule rdfs:label ?chembl\_id .

 ?act cco:hasAssay ?ass.

 ?act cco:standardValue ?std\_value .

 ?act cco:standardUnits ?std\_unit .

 ?act cco:standardType ?std\_type .

 ?act cco:standardRelation ?std\_rel.

 ?ass dcterms:description ?ass\_disc.

 ?ass cco:hasTarget ?tar.

 ?tar rdfs:label ?tar\_id .}

?s  cco:hastarget ?tar.

?tar cco:hasTargetComponent  ?tarcom

?tarcom cco:hastarget ?a

!!!!!!!! hastarget twice

**Retrieve compounds which are selective to one target over a second target:**

SELECT  ?chembl\_id  ?std\_value ?std\_unit  ?std\_type

WHERE {

?molecule cco:hasActivity ?Activity.

?molecule rdfs:subClassOf cco:Substance.

?Activity cco:standardValue ?std\_value .

?Activity cco:standardUnits ?std\_unit .

?Activity cco:standardType ?std\_type .

?molecule rdfs:label ?chembl\_id .

FILTER ( regex(?std\_value < 50.0) && regex(?std\_unit ,"nM" )&& regex(?std\_type, "IC50")   ).

}

SELECT  DISTINCT ?molecule ?chembl\_id  ?std\_value ?std\_unit  ?std\_type

WHERE {

?molecule cco:hasActivity ?Activity.

?molecule rdfs:subClassOf cco:Substance.

?Activity cco:standardValue ?std\_value .

?Activity cco:standardUnits ?std\_unit .

?Activity cco:standardType ?std\_type .

?Acctivity cco:hasAssay ?ass.

?ass cco:hasTarget ?tar.

?molecule rdfs:label ?chembl\_id .

FILTER ( (?std\_value >= 200  && regex(?std\_unit ,"nM" )&& regex(?std\_type, "IC50") && regex( ?tar , "CHEMBL4036")) )

}

SELECT  DISTINCT ?molecule ?chembl\_id  ?std\_value ?std\_unit  ?std\_type

WHERE {

?molecule cco:hasActivity ?Activity.

?molecule rdfs:subClassOf cco:Substance.

?Activity cco:standardValue ?std\_value .

?Activity cco:standardUnits ?std\_unit .

?Activity cco:standardType ?std\_type .

?Acctivity cco:hasAssay ?ass.

?ass cco:hasTarget ?tar.

?molecule rdfs:label ?chembl\_id .

FILTER ( (?std\_value >= 200  && regex(?std\_unit ,"nM" )&& regex(?std\_type, "IC50") && regex( ?tar , "CHEMBL4036")) )

}

INTERSECT

SELECT  DISTINCT ?molecule ?chembl\_id  ?std\_value ?std\_unit  ?std\_type

WHERE {

?molecule cco:hasActivity ?Activity.

?molecule rdfs:subClassOf cco:Substance.

?Activity cco:standardValue ?std\_value .

?Activity cco:standardUnits ?std\_unit .

?Activity cco:standardType ?std\_type .

?Acctivity cco:hasAssay ?ass.

?ass cco:hasTarget ?tar.

?molecule rdfs:label ?chembl\_id .

FILTER ( (?std\_value < 50.0) && regex(?std\_unit ,"nM" )&& regex(?std\_type, "IC50") && regex( ?tar , "CHEMBL4036"))  ).

}

#**Retrieve target ChEMBL\_ID, target\_name, target\_type, protein accessions and sequences for all protein targets:**

PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>

PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>

PREFIX owl: <http://www.w3.org/2002/07/owl#>

PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>

PREFIX dc: <http://purl.org/dc/elements/1.1/>

PREFIX dcterms: <http://purl.org/dc/terms/>

PREFIX dbpedia2: <http://dbpedia.org/property/>

PREFIX dbpedia: <http://dbpedia.org/>

PREFIX foaf: <http://xmlns.com/foaf/0.1/>

PREFIX skos: <http://www.w3.org/2004/02/skos/core#>

PREFIX cco: <http://rdf.ebi.ac.uk/terms/chembl#>

SELECT ?tar ?title ?tarTyp  ?protSeq ?uniref

WHERE {

 ?o cco:hasTarget ?tar.

 ?tar dcterms:title ?title.

 ?tar cco:targetType ?tarTyp.

 ?tar cco:hasTargetComponent ?tc.

 ?tc cco:hasProteinClassification ?pcl.

 ?tc cco:proteinSequence ?protSeq.

 ?tc skos:exactMatch ?em.

?em rdfs:label ?uniref.

}

#**Retrieve compound activity details for all targets containing a protein of interest:**

#Protein of interest is human M2 muscarinic receptor (P08172)

PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>

PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>

PREFIX owl: <http://www.w3.org/2002/07/owl#>

PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>

PREFIX dc: <http://purl.org/dc/elements/1.1/>

PREFIX dcterms: <http://purl.org/dc/terms/>

PREFIX dbpedia2: <http://dbpedia.org/property/>

PREFIX dbpedia: <http://dbpedia.org/>

PREFIX foaf: <http://xmlns.com/foaf/0.1/>

PREFIX skos: <http://www.w3.org/2004/02/skos/core#>

PREFIX cco: <http://rdf.ebi.ac.uk/terms/chembl#>

PREFIX bibo: <http://purl.org/ontology/bibo/>

SELECT ?chembl\_id ?std\_value ?std\_unit ?pub\_id ?std\_type  ?assy\_description ?tar ?title ?tarTyp

WHERE {

 ?molecule cco:hasActivity ?act.

 ?act cco:hasDocument ?doc.

 OPTIONAL{?act cco:standardValue ?std\_value .}

 OPTIONAL{?act cco:standardUnits ?std\_unit .}

 OPTIONAL{?act cco:standardType ?std\_type .}

 ?doc bibo:pmid ?pub\_id.

 ?molecule rdfs:subClassOf cco:Substance.

 ?molecule rdfs:label ?chembl\_id.

 ?act cco:hasAssay ?ass.

 ?ass dcterms:description ?assy\_description.

 ?ass cco:hasTarget ?tar.

 ?tar dcterms:title ?title.

 ?tar cco:targetType ?tarTyp.

 ?tar cco:hasTargetComponent ?tc.

 ?tc skos:exactMatch ?em.

 ?em rdfs:label ?uniref.

 FILTER regex(?uniref ,"P08172")

}

# “CHEMBL282282” compound publication journal and date

PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>

PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>

PREFIX owl: <http://www.w3.org/2002/07/owl#>

PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>

PREFIX dc: <http://purl.org/dc/elements/1.1/>

PREFIX dcterms: <http://purl.org/dc/terms/>

PREFIX dbpedia2: <http://dbpedia.org/property/>

PREFIX dbpedia: <http://dbpedia.org/>

PREFIX foaf: <http://xmlns.com/foaf/0.1/>

PREFIX skos: <http://www.w3.org/2004/02/skos/core#>

PREFIX cco: <http://rdf.ebi.ac.uk/terms/chembl#>

PREFIX bibo: <http://purl.org/ontology/bibo/>

SELECT DISTINCT ?journal ?title ?date

WHERE{

 ?molecule rdfs:subClassOf cco:Substance.

 ?molecule rdfs:label "CHEMBL282282".

 ?molecule cco:hasDocument ?doc.

  ?doc cco:hasJournal ?journal.

  ?journal dcterms:title ?title.

  ?doc dcterms:date ?date.

}

#LIST Target Type Available in Database:

PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>

PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>

PREFIX owl: <http://www.w3.org/2002/07/owl#>

PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>

PREFIX dc: <http://purl.org/dc/elements/1.1/>

PREFIX dcterms: <http://purl.org/dc/terms/>

PREFIX dbpedia2: <http://dbpedia.org/property/>

PREFIX dbpedia: <http://dbpedia.org/>

PREFIX foaf: <http://xmlns.com/foaf/0.1/>

PREFIX skos: <http://www.w3.org/2004/02/skos/core#>

PREFIX cco: <http://rdf.ebi.ac.uk/terms/chembl#>

PREFIX bibo: <http://purl.org/ontology/bibo/>

SELECT DISTINCT ?tarTyp

WHERE {

 ?ass cco:hasTarget ?tar.

 ?tar cco:targetType ?tarTyp.

}

#Compound targeting “Firefly Luciferase”

PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>

PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>

PREFIX owl: <http://www.w3.org/2002/07/owl#>

PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>

PREFIX dc: <http://purl.org/dc/elements/1.1/>

PREFIX dcterms: <http://purl.org/dc/terms/>

PREFIX dbpedia2: <http://dbpedia.org/property/>

PREFIX dbpedia: <http://dbpedia.org/>

PREFIX foaf: <http://xmlns.com/foaf/0.1/>

PREFIX skos: <http://www.w3.org/2004/02/skos/core#>

PREFIX cco: <http://rdf.ebi.ac.uk/terms/chembl#>

SELECT DISTINCT ?chembl\_id ?label

WHERE {

 ?molecule rdfs:subClassOf cco:Substance.

 ?molecule rdfs:label ?chembl\_id .

 ?molecule cco:hasActivity ?act.

 ?act cco:hasAssay ?ass.

 ?ass cco:hasTarget ?tar.

 ?tar rdfs:label ?label.

 ?tar dcterms:title "Firefly luciferase".

}

# subtance type interaction with cell-lines.

PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>

PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>

PREFIX owl: <http://www.w3.org/2002/07/owl#>

PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>

PREFIX dc: <http://purl.org/dc/elements/1.1/>

PREFIX dcterms: <http://purl.org/dc/terms/>

PREFIX dbpedia2: <http://dbpedia.org/property/>

PREFIX dbpedia: <http://dbpedia.org/>

PREFIX foaf: <http://xmlns.com/foaf/0.1/>

PREFIX skos: <http://www.w3.org/2004/02/skos/core#>

PREFIX cco: <http://rdf.ebi.ac.uk/terms/chembl#>

PREFIX sio: <http://semanticscience.org/resource/>

SELECT DISTINCT   ?m  ?label

WHERE {

      ?molecule rdfs:subClassOf ?subs;

           cco:substanceType ?m;

           skos:prefLabel ?prlabel ;

           rdfs:label ?chembl\_id ;

           cco:hasActivity ?act.

 ?act cco:hasAssay ?ass.

 ?ass cco:hasTarget ?tar.

 ?tar cco:targetType "CELL-LINE".

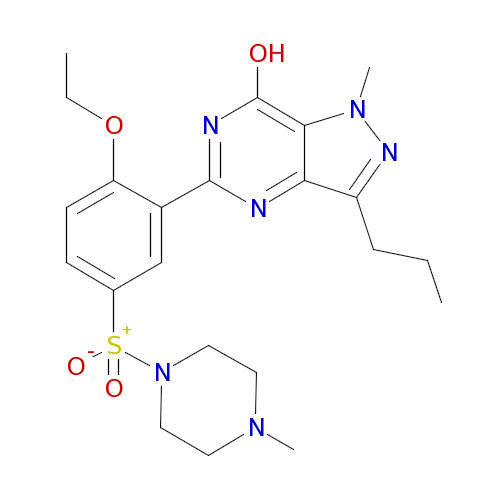
 ?tar rdfs:label ?label.

}

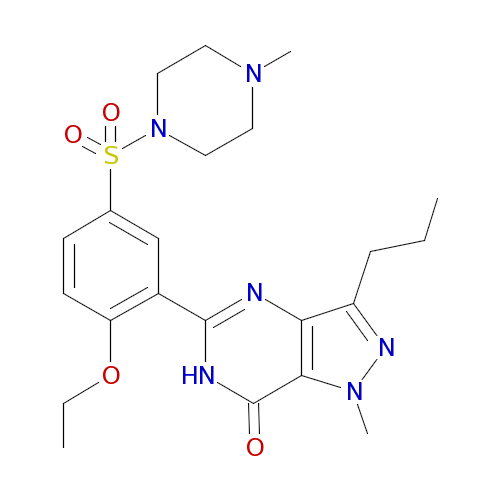
TARGET TYPE

|  |
| --- |
| ORGANISM |
| SINGLE PROTEIN |
| PROTEIN FAMILY |
| UNCHECKED |
| PROTEIN COMPLEX |
| ADMET |
| CELL-LINE |
| NUCLEIC-ACID |
| PHENOTYPE |
| PROTEIN COMPLEX GROUP |
| SELECTIVITY GROUP |
| UNKNOWN |
| TISSUE |
| PROTEIN-PROTEIN INTERACTION |
| SUBCELLULAR |
| NO TARGET |
| CHIMERIC PROTEIN |
| SMALL MOLECULE |
| MACROMOLECULE |

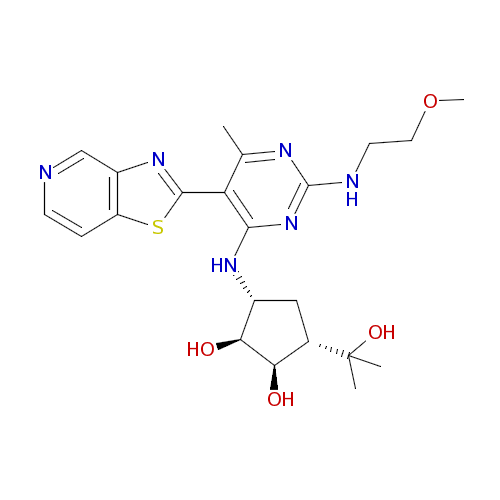
|  |
| --- |
| subsTyp |
| Cell |
| Small molecule |
| Enzyme |
| Unclassified |
| Unknown |
| Protein |
| Oligonucleotide |
| Antibody |
| Oligosaccharide |



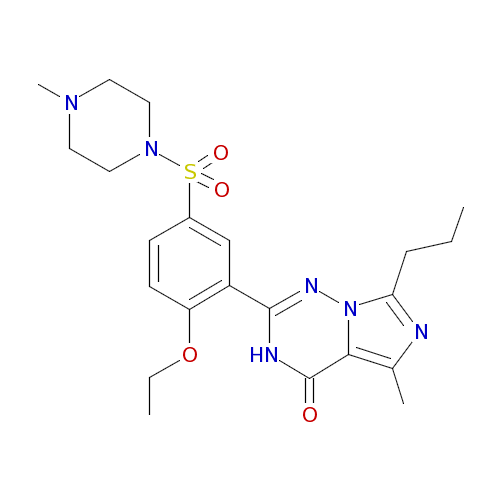
1886521



192



2163755



168949

|  |
| --- |
|  |
|  |

169843

|  |
| --- |
|  |

183370

|  |
| --- |
|  |

2331679

|  |
| --- |
|  |

2181648