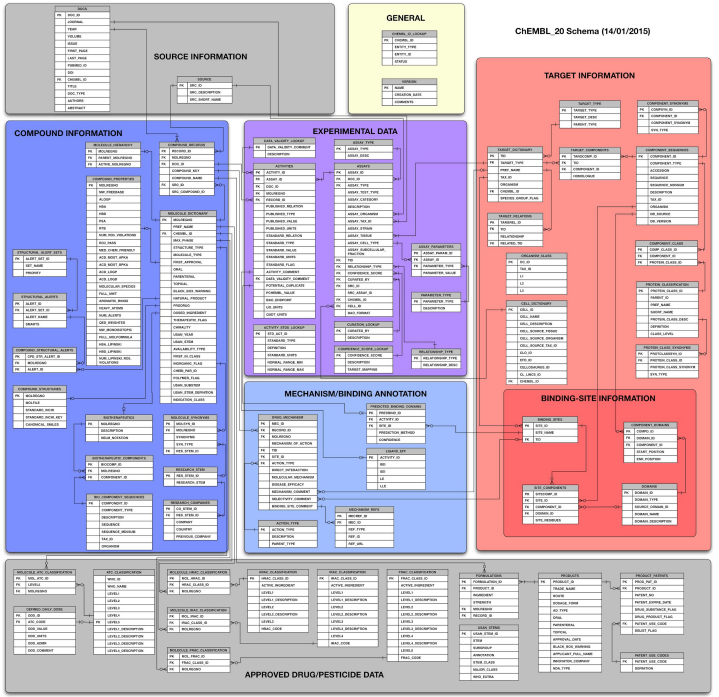
ChemAxon

János Kendi

Chemical Search in Graph Database

Why graph database?

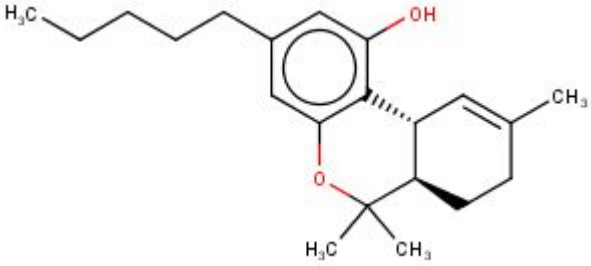
- Relationships > Structural data
- *SELECT TableA.*, TableB.*, TableC.*, TableD.*
  FROM TableA
  JOIN TableB
  ON TableB.aID = TableA.aID
  JOIN TableC
  ON TableC.cID = TableB.cID
  JOIN TableD
  ON TableD.dID = TableA.dID
  WHERE ...

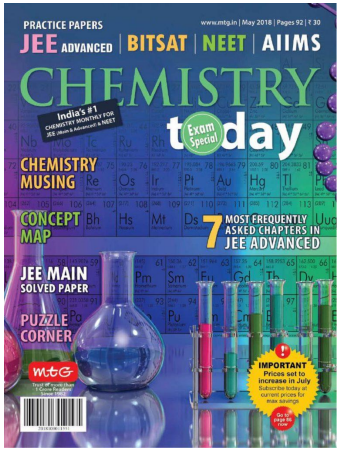


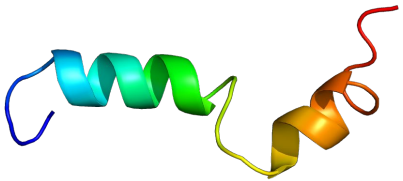
ChemAxon

Nodes

- (molecule)
- (target)
- (document)







ChemAxon

Relations

- (molecule)-[:ACTIVE_ON]->(target)
- (molecule)-[:MENTIONED_IN]->(document)
- (target)-[:MENTIONED_IN]->(document)
- **(molecule)-[:SIM_90]->(molecule)**

Similarity

```
match (n:molecule)
call jchem.search('neo4j_demo', n.smilesString, 0, 'sim', 0.9) yield node
where n<>node
create (n)-[r:SIM_90]->(node)
```

Demo

Find an interesting compound

```
call jchem.search('neo4j_demo', 'LSD', 1, 'dup') yield node
return node;
```

Check activity (in homo sapiens)

```
call jchem.search('neo4j_demo','LSD', 1, 'dup') yield node
with node
match (node)-[activity:ACTIVE_ON]->(t:target {organism: 'Homo sapiens'})
where activity.pchembl >= 7
return t;
```

Find other active compounds on these targets

```
call jchem.search('neo4j_demo','LSD', 1, 'dup') yield node
with node
match (node)-[activity:ACTIVE_ON]->(t:target {organism:
'Homo sapiens'})<-[activity2:ACTIVE_ON]-(mol:molecule)
where activity.pchembl >= 7 and activity2.pchembl >= 7
return mol limit 100;
```

Too many, filter the similar ones

```
call jchem.search('neo4j_demo','LSD', 1, 'dup') yield node
with node
match (node)-[activity:ACTIVE_ON]->(t:target {organism:
'Homo sapiens'})
<-[activity2:ACTIVE_ON]-(mol:molecule)-[:SIM_90]->(node)
where activity.pchembl >= 7 and activity2.pchembl >= 7
return mol;
```

More info about these compounds

```
call jchem.search('neo4j_demo','LSD', 1, 'dup') yield node
with node
match (node)-[activity:ACTIVE_ON]->(t:target {organism:
'Homo sapiens'})
<-[activity2:ACTIVE_ON]-(mol:molecule)-[:SIM_90]->(node),
(mol)-[:MENTIONED_IN]->(doc:document)
where activity.pchembl >= 7 and activity2.pchembl >= 7
return doc;
```

ChemAxon

All in one

```
call jchem.search('neo4j_demo','LSD', 1, 'dup') yield node
with node
match (node)-[activity:ACTIVE_ON]->(t:target {organism:
'Homo sapiens'})
<-[activity2:ACTIVE_ON]-(mol:molecule)-[:SIM_90]->(node),
(mol)-[r:MENTIONED_IN]->(doc:document)
where activity.pchembl >= 7 and activity2.pchembl >= 7
return node, mol, activity, activity2, t, r, doc;
```

THANK YOU

ChemAxon