Drug discovery FAQs: workflows for answering multidomain drug discovery questions

Daniela Digles
University of Vienna
support@openphacts.org

4th KNIME Cheminformatics Workshop
Data collection
Useful links

- Latest version of the API: https://dev.openphacts.org/docs/1.5
- Support portal: http://support.openphacts.org/
- Open PHACTS Explorer: http://explorer.openphacts.org/
- Example Workflows: http://www.myexperiment.org/groups/1125.html
Answering „scientific competency questions“

- 20 questions defined at the beginning of the project.
- Two main clusters:
  1.) Compound-target relationships
  2.) includes additional complexity of diseases, pathways, text-mining and patents.

- Example: Give me all oxidoreductase inhibitors active <100 nM in human and mouse.
- Many questions need a combination of queries to the Open PHACTS Platform.
## Cluster 1 use cases (Q1-Q11): Answers require mainly compound-target pharmacology data

<table>
<thead>
<tr>
<th>ID</th>
<th>Use case question</th>
<th>Sequence of API calls with filters used and link to the workflow</th>
</tr>
</thead>
</table>
| Q1 | Give me all oxidoreductase inhibitors active <100 nM in human and mouse. | **Target Class Pharmacology** (target_organism=Homo sapiens|Mus musculus; minEx-pChembl=7)  
http://www.myexperiment.org/workflows/4504.html |
| Q2 | For a given compound what is its predicted secondary pharmacology? | **Compound Information**> **Chemical Structure Search: Similarity**> **Compound adverse events** |
| Q3 | Given a target find me all actives against that target. Find/predict polypharmacology of actives. | **Target Pharmacology** (minEx-pChembl=5) > **Compound Pharmacology** (minEx-pChembl=0)  
http://www.myexperiment.org/workflows/4505.html |
| Q4 | For a given interaction profile, give me similar compounds. | **Compound Information** > **Compound Information (Batch)** > **Chemical Structure Search: Similarity** (searchOptions.Threshold=0.85)> **Compound Information**  
http://www.myexperiment.org/workflows/4516.html |
| Q5 | For molecules that contain substructure X, retrieve all bioactivity data in serine protease assays. | **Chemical Structure Search: Substructure**> **Compound Pharmacology**, **Target Class Members**  
http://www.myexperiment.org/workflows/4478.html |
| Q6 | For a specific target family, retrieve all compounds in specific assays | **Target Class Pharmacology**  
http://www.myexperiment.org/workflows/4506.html |
<p>| Q7 | For a target, give me all active compounds | <strong>Target Pharmacology</strong> (minEx-pChembl=5) |</p>
<table>
<thead>
<tr>
<th><strong>ID</strong></th>
<th><strong>Use case question</strong></th>
<th><strong>Sequence of API calls with filters used and link to the workflow</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>Q6</td>
<td>For a specific target family, retrieve all compounds in specific assays</td>
<td><strong>Target Class Pharmacology</strong></td>
</tr>
<tr>
<td></td>
<td></td>
<td><a href="http://www.myexperiment.org/workflows/4506.html">http://www.myexperiment.org/workflows/4506.html</a></td>
</tr>
<tr>
<td>Q7</td>
<td>For a target, give me all active compounds with the relevant assay data.</td>
<td><strong>Target Pharmacology</strong> <em>(minEx-pChembl=5)</em></td>
</tr>
<tr>
<td></td>
<td></td>
<td><a href="http://www.myexperiment.org/workflows/4507.html">http://www.myexperiment.org/workflows/4507.html</a></td>
</tr>
<tr>
<td>Q8</td>
<td>Identify all known protein-protein interaction inhibitors</td>
<td><strong>Target Class Pharmacology</strong> <em>(target_type=ppi, minEx-pChembl=5)</em></td>
</tr>
<tr>
<td></td>
<td></td>
<td><a href="http://www.myexperiment.org/workflows/4508.html">http://www.myexperiment.org/workflows/4508.html</a></td>
</tr>
<tr>
<td>Q9</td>
<td>For a given compound, give me the interaction profile with targets.</td>
<td><strong>Compound Pharmacology</strong> *(activity_type=IC50</td>
</tr>
<tr>
<td></td>
<td></td>
<td><a href="http://www.myexperiment.org/workflows/4509.html">http://www.myexperiment.org/workflows/4509.html</a></td>
</tr>
<tr>
<td>Q10</td>
<td>For a given compound, summarize all similar compounds and their activities.</td>
<td><strong>Chemical Structure Search: Similarity</strong> <em>(searchOptions.SimilarityType=0; searchOptions.Threshold=0.80)</em>&gt; <strong>Compound Pharmacology</strong> *(activity_type=IC50</td>
</tr>
<tr>
<td></td>
<td></td>
<td><a href="http://www.myexperiment.org/workflows/4510.html">http://www.myexperiment.org/workflows/4510.html</a></td>
</tr>
<tr>
<td>Q11</td>
<td>Retrieve all data for a given list of compounds depicted by their chemical structure (SMILES) with options to match stereochemistry.</td>
<td><strong>Chemical Structure Search: Exact</strong> <em>(searchOptions.MatchType=2)</em>&gt; <strong>Compound Pharmacology, Compound Information, Compound Classifications</strong> <em>(tree=chebi)</em></td>
</tr>
<tr>
<td></td>
<td></td>
<td><a href="http://www.myexperiment.org/workflows/4511.html">http://www.myexperiment.org/workflows/4511.html</a></td>
</tr>
<tr>
<td>Q12</td>
<td>For a given compound, which of its targets have been patented in the context of a disease?</td>
<td><strong>Compound Pharmacology &gt; PatentsCalls &gt; Disease for Target</strong></td>
</tr>
<tr>
<td>Q13</td>
<td>For disease X, which targets have ligands in different stages of the development process with publications/patents describing these compounds?</td>
<td><strong>Targets for Disease &gt; Target Pharmacology (minEx-pChembl=5), Target Information &gt; Patents calls</strong></td>
</tr>
<tr>
<td>Q14</td>
<td>Target druggability: compounds directed against target X have what indications? Which new targets have appeared recently in the patent literature for a disease?</td>
<td><strong>Target pharmacology (minEx-pChembl=5) &gt; Indications for Compounds &gt; Patent calls &gt; Disease for Targets</strong></td>
</tr>
</tbody>
</table>
| Q15 | a) Which chemical series have been shown to be active against target X? b) Which new targets have been associated with disease Y? c) Which companies are working on target X or disease Y? | a) **Classification of Compounds for Target (minEx-pChembl=5)**  
b) **Associations for Disease**  
c) Competitive Intelligence data not available  
http://www.myexperiment.org/workflows/4512.html |
| Q16 | Targets in Parkinson's disease or Alzheimer's disease are activated by which compounds? | **Target for Disease > Target Pharmacology (minEx-pChembl=5)**  
http://www.myexperiment.org/workflows/4513.html |
| Q17 | For my specific target, which active compounds have been reported in the literature? | **Target Pharmacology (minEx-pChembl=5)**  
http://www.myexperiment.org/workflows/4507.html |
| Q18 | For pathway X, find compounds that agonize targets assayed in only functional assays with potency <1 µM. | **Pathway Information: Get Targets > Target Pharmacology**  
(activity_type=Potency, max-activity_value=1000, activity_unit=nanomolar)  
http://www.myexperiment.org/workflows/4514.html |
| Q19 | For the targets in a given pathway, retrieve the compounds that are active with more than one target. | **Pathway Information: Get Targets > Target Pharmacology (minEx-pChembl=5)**  
http://www.myexperiment.org/workflows/4515.html |
| Q20 | For a given disease, retrieve all targets in the pathway and all active compounds hitting them. | **Targets for Disease > Target Pharmacology (minEx-pChembl=5)**  
http://www.myexperiment.org/workflows/4513.html |
Example workflow

- Q10: For a given compound, summarize all similar compounds and their activities

```
CC1=C(C(C(=C(N1)C)C(=O)OC)C2
=CC=CC=C2[N+](=O)[O-])C(=O)OC
```

![Diagram of the workflow process](image-url)
OPS-Knime nodes

- Created by Ronald Siebes, VU Amsterdam.
- No predefined set of nodes for each API call.

- **OPS_Swagger:**
  - creates the API call
  - Swagger file is used to automatically provide available API calls and parameters

- **OPS_JSON (deprecated):**
  - executes the API call
  - transforms the output into a flattened spreadsheet format

- available from https://github.com/openphacts/OPS-Knime
Swagger

- Structured format for the generation of API documentation. (https://helloreverb.com/developers/swagger)
- https://raw.githubusercontent.com/openphacts/OPS_LinkedDataApi/1.5.0/api-config-files/swagger.json

```json
{
  "basePath": "https://beta.openphacts.org/1.3",
  "apiVersion": "v1.3",
  "apis": [
    {
      "path": "/compound",
      "operations": [
        {
          "httpMethod": "GET",
          "summary": "Compound Information",
          "description": "Information about a single compound.

          "group": "Compound",
          "parameters": [
            {
              "name": "uri",
              "description": "A compound URI. e.g.: http://www.conceptwiki.org/concept/38932552-111f-4a4e-a46a-4edfbf4b1357",
              "dataType": "string",
              "required": true,
              "paramType": "query"
            },
            ...
        }
      ]
    }
  ]
}
```
OPS_Swagger details

- Knime node where the user provides a url to a Swagger file (default: Open PHACTS API, v1.4)
- File is parsed and provides a list of the available API calls.
- Parameters tab is updated to the available parameters.
- Parameters can be set in the parameters tab or in the input table.
- Output of the node is an executable API call.
OpenPHACTS API v1.5

**Compound Information**

**Description**
Returns information about a single compound including (but not limited to): molecular weight, biotransformation, protein binding and toxicity.

<table>
<thead>
<tr>
<th>PARAMETER</th>
<th>VALUE</th>
<th>DESCRIPTION</th>
</tr>
</thead>
<tbody>
<tr>
<td>uri</td>
<td>(required)</td>
<td>A compound URL e.g.: <a href="http://www.conceptwiki.org/concept/38932552-111f-4a4e-a46a-4ed1d7bdf9d5">http://www.conceptwiki.org/concept/38932552-111f-4a4e-a46a-4ed1d7bdf9d5</a></td>
</tr>
<tr>
<td>app_id</td>
<td></td>
<td>Your access application id</td>
</tr>
<tr>
<td>app_key</td>
<td></td>
<td>Your access application key</td>
</tr>
<tr>
<td>_format</td>
<td></td>
<td>The desired result format.</td>
</tr>
</tbody>
</table>
OPS_Swagger details

![Screen capture of OPS_Swagger with selected options and URL]

- **Swagger URL:** `linkedDataApi/1.5.0/api-config-files/swagger.js`
OPS_Swagger details

![Dialog - 3:126:40 - OPS_Swagger(Targets for Disease)](image)

- **uri**: [Blank]
- **app_id**: 15a18100
- **app_key**: 272f1cd961d215f318a0315dd3d
- **_page**: [Blank]
- **_pageSize**: all
- **_orderBy**: [Blank]
- **_format**: [Blank]
- **_callback**: [Blank]
- **_metadata**: [Blank]
OPS_Swagger details

executable API call
Retrieving and parsing the results

- Either use OPS_Json (deprecated) or the REST and JSON nodes available as addin from KNIME.
- GET Resource: retrieves the actual data from the server. Configure the node to use the column url as input. Response representation cell type: Autodetection.
- String to JSON: transforms the result to a JSON column type.
- JSON Path: allows the individual selection of the data which is transformed into a tabular structure.
Example workflow

Q10: For a given compound, summarize all similar compounds and their activities

CC1=C(C(C(=C(N1)C)C(=O)OC)C2=CC=CC=C2[N+]1(O)[O-])C(=O)OC
Similarity Search Metanode
Compound Pharmacology Metanode
Difficulties

- Definition of „active“ compounds
- Wish of retrieving „all“ results
- Requested data not available from open access data providers (e.g. Which companies are working on target X or disease Y?)
- Data not yet available in the Open PHACTS Discovery Platform (Patents; upstream/downstream information from Pathways)
- Long execution times
Current activities

- Improvement of the available Open PHACTS Knime nodes
- Improvement of the existing workflows:
  - Replacement of OPS_JSON node with existing nodes available from Knime
  - Update of the API to the latest available version
- Integration of new data sources (e.g. patent data)
Acknowledgements

- Pharmacoinformatics research group, University of Vienna
  - Gerhard F. Ecker
  - Barbara Zdrazil
  - Lars Richter

- Open PHACTS – KNIME
  - Ronald Siebes, VU Amsterdam
  - Christine Chichester, SIB
  - Evan Tzanis, QMUL
GENOMICS England & OPENPHACTS

Evan Tzanis

QMUL
Outline

- Project Definition
- Engineering Approach
- KNIME related questions
Queries

– What small molecules act on the following genes and are orally bioavailable (based on lipinsky’s rule of 5)

– What molecules act on the following genes and are likely to be brain penetrant

– What molecules acting on the following genes have side effects that correlate with a specific phenotype?
Answering Business questions with pipelines

- For molecules that contain substructure X, retrieve all bioactivity data in serine protease assays
- KNIME or Pipeline Pilot compatible
  - pre-configured node library
  - eg. R, python, NGS tools
- Enable API linkage and pipelining

KNIME Questions

• Which is the current trend in KNIME for developing nodes against an API?

• Code Requirements for a Node being part of KNIME's Node collection.

• Hints for developing nice Front Ends for KNIME Nodes.
KNIME Questions

- KNIME Nodes for Reporting.
- KNIME nodes for storing data to external resources.
- Recommended KNIME debugging tools.