Collecting compound data from public databases using Open PHACTS and KNIME

Daniela Digles and Gerhard F. Ecker
University of Vienna, Department of Medicinal Chemistry, Althanstraße 14, 1090, Vienna

Introduction

The Open PHACTS project [1] integrates several public databases, thus allowing answering questions relevant for research in the drug-discovery process [2]. The data collected in the project can be accessed with web tools, such as the Open PHACTS Explorer (www.openphacts.org/explorer) or by using the programming interface (API, available from dev.openphacts.org). To answer questions requiring several different queries in an automated way, workflow tools such as KNIME or Pipeline Pilot are a valuable resource. Such KNIME workflows were used for several use-cases [3] and to provide answers to drug-discovery research questions [4].

Aim of the work

In the current work, a workflow to easily generate a summary of known data connected to a compound of interest is presented.

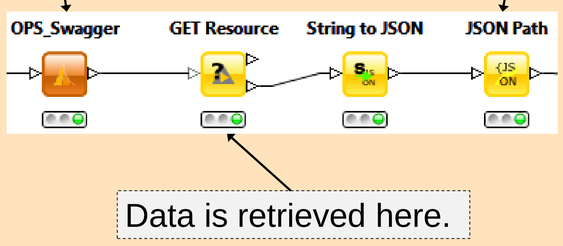
Methods

The workflow was generated using KNIME, with the Open PHACTS nodes available from the github repository (https://github.com/openphacts/OPS-Knime).

General implementation of API calls with KNIME nodes

OPS_Swagger generates the API call.

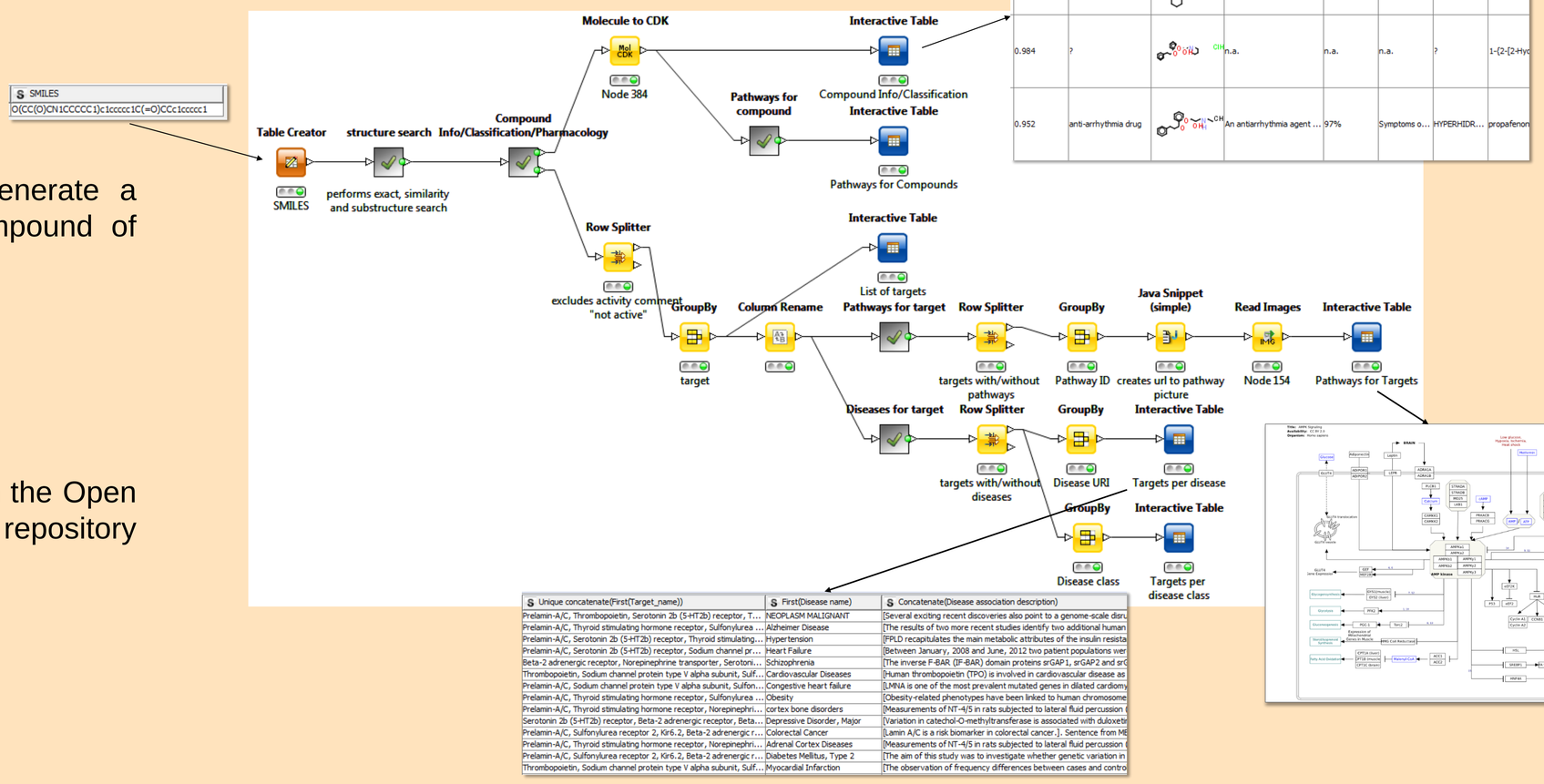
Columns which should be included in the tabular output are specified here.



Currently, a multi-step approach is necessary to retrieve data from the API. A simplified node to retrieve the data directly is currently under development.

The workflow

- Query: SMILES of the molecule of interest.
- Exact or similarity/substructure search.
- Data for retrieved molecules:
  - Function and toxicity annotation (Drugbank)
  - Role of the molecule (ChEBI)
  - Pharmacology data, pChEMBL >= 5 (ChEMBL)
- Data for retrieved targets:
  - Pathways (WikiPathways)
  - Diseases (DisGeNET)
- Simple counts generated by GroupBy node.



Unique concatenate(First(Target_name))	First(Disease name)	Concatenate(Disease association description)
Prelamin-A/C, Thrombopoietin, Serotonin 2b (5-HT2b) receptor, T…	NEOPLASM MALIGNANT	[Several exciting recent discoveries also point to a genome-scale disru
Prelamin-A/C, Thyroid stimulating hormone receptor, Sulfonylurea …	Alzheimer Disease	[The results of two more recent studies identify two additional human
Prelamin-A/C, Serotonin 2b (5-HT2b) receptor, Thyroid stimulating…	Hypertension	[FPLD recapitulates the main metabolic attributes of the insulin resista
Prelamin-A/C, Serotonin 2b (5-HT2b) receptor, Sodium channel pr…	Heart Failure	[Between January, 2008 and June, 2012 two patient populations wer
Beta-2 adrenergic receptor, Norepinephrine transporter, Serotoni…	Schizophrenia	[The inverse F-BAR (IF-BAR) domain proteins srGAP1, srGAP2 and srG
Thrombopoietin, Sodium channel protein type V alpha subunit, Sulf…	Cardiovascular Diseases	[Human thrombopoietin (TPO) is involved in cardiovascular disease as
Prelamin-A/C, Sodium channel protein type V alpha subunit, Sulfon…	Congestive heart failure	[LMNA is one of the most prevalent mutated genes in dilated cardiomy
Prelamin-A/C, Thyroid stimulating hormone receptor, Sulfonylurea …	Obesity	[Obesity-related phenotypes have been linked to human chromosome
Prelamin-A/C, Thyroid stimulating hormone receptor, Norepinephri…	cortex bone disorders	[Measurements of NT-4/5 in rats subjected to lateral fluid percussion (
Serotonin 2b (5-HT2b) receptor, Beta-2 adrenergic receptor, Beta…	Depressive Disorder, Major	[Variation in catechol-O-methyltransferase is associated with duloxetir
Prelamin-A/C, Sulfonylurea receptor 2, Kir6.2, Beta-2 adrenergic r…	Colorectal Cancer	[Lamin A/C is a risk biomarker in colorectal cancer.]. Sentence from MB
Prelamin-A/C, Thyroid stimulating hormone receptor, Norepinephri…	Adrenal Cortex Diseases	[Measurements of NT-4/5 in rats subjected to lateral fluid percussion (
Prelamin-A/C, Sulfonylurea receptor 2, Kir6.2, Beta-2 adrenergic r…	Diabetes Mellitus, Type 2	[The aim of this study was to investigate whether genetic variation in
Thrombopoietin, Sodium channel protein type V alpha subunit, Sulf…	Myocardial Infarction	[The observation of frequency differences between cases and contro

Results

Example: O(CC(O)CN1CCCCC1)c1ccccc1C(=O)CCc1ccccc1
(a propafenone derivative)

- Structure search: 96 molecules (including the query)
- 1 known drug (propafenone)
- 191 activity values (< 10µM) against 33 targets
- 98 pathways
- >2000 diseases in 25 disease classes

Conclusions

The workflow allows the preparation of a first overview on data known for a potential compound of interest, including data for similar compounds. Links to the original data sources are retained, allowing manual curation of the collected associations.

Acknowledgement:
The research leading to these results has received support from the Innovative Medicines Initiative Joint Undertaking under grant agreement no. [115191], resources of which are composed of financial contribution from the European Union's Seventh Framework Programme (FP7/2007-2013) and in-kind contribution of EFPIA companies.

References:
[1] Williams A J, Harland L, Groth P, *et al.* (2012) Drug Discov. Today 17: 1188 – 1198.
[2] Azzaoui K, Jacoby E, Senger S, *et al.* (2013) Drug Discov. Today 18: 843 – 852.
[3] Ratnam J, Zdrazil B, Digles D, *et al.* (2014) PloS One: e115460. doi: 10.1371/journal.pone.0115460.
[4] Chichester C, Digles D, Siebes R, *et al.* (2015) Drug Discov. Today 20: 399 – 405.