Collecting compound data from public databases using Open PHACTS and KNIME

Daniela Digles, Gerhard F. Ecker

University of Vienna, Department of Pharmaceutical Chemistry, Althanstraße 14, 1090 Vienna, Austria.

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].

In the current work, a workflow to easily generate a summary of known data connected to a compound of interest is presented. The workflow was generated using KNIME, with the Open PHACTS nodes available from https://github.com/openphacts/OPS-Knime.

Starting from the depiction of a molecule, similar molecules are retrieved. Information collected directly for the molecule includes function and toxicity annotations (from Drugbank), the role of the molecule (from ChEBI), and biological pathways containing this molecule (from Wikipathways). In the next step, proteins where the molecule is reported to be active in ChEMBL are returned, and the connection of the proteins to biological pathways (from Wikipathways) and to diseases (from DisGeNET) are shown. The data from all these sources was retrieved via the Open PHACTS API, easily connecting the identifiers used in the different databases. The finished workflow can be easily adapted to query for a different molecule, and can be executed without knowledge about programming.

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.

Acknowledgements: The work 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.