Using linked open data for assessing multi-target SAR

Barbara Zdrazil,[1] Daniela Digles,[1] Edgar Jacoby,[2] Stefan Senger,[3] Gerhard F. Ecker[1]

1 University of Vienna, Department of Pharmaceutical Chemistry, Vienna, Austria
2 Janssen Research & Development, Beerse, Belgium
3 GSK Medicines Research Centre, Stevenage, Hertfordshire, United Kingdom

Email to: barbara.zdrazil@univie.ac.at

References

[1] Ratnam J, Zdrazil B, Digles D, Cuadrado-Rodriguez E, Neefs JM, et al. (2014) The Application of the Open Pharmacological Concepts Triple Store (Open PHACTS) to Support Drug Discovery Research. PLOS ONE 9(12): e115460
[2] Williams AJ, Harland L, Groth P, Pettifer S, Chichester C, et al. (2012) Open PHACTS: Semantic interoperability for drug discovery. Drug Discov Today 17:1188–1198. doi: 10.1016/j.drudis.2012.05.016
[3] Berthold M, Cebron N, Dill F, Gabriel T, Kötter T, et al.. (2008) KNIME: The konstanz information miner. In: Preisach C, Burkhardt H, Schmidt-Thieme L, Decker R, editors.: Springer Berlin Heidelberg. pp.319–326.; http://www.knime.org
[4] Bemis GW, Murcko MA (1996) The Properties of Known Drugs. 1. Molecular Frameworks. J Med Chem 39: 2887–2893.

The research leading to these results has received support from the Innovative Medicines Initiative Joint Undertaking under grant agreement n° 115191 (Open PHACTS), resources of which are composed of financial contribution from the European Union's Seventh Framework Programme (FP7/2007-2013) and EFPIA companies' in kind contribution.

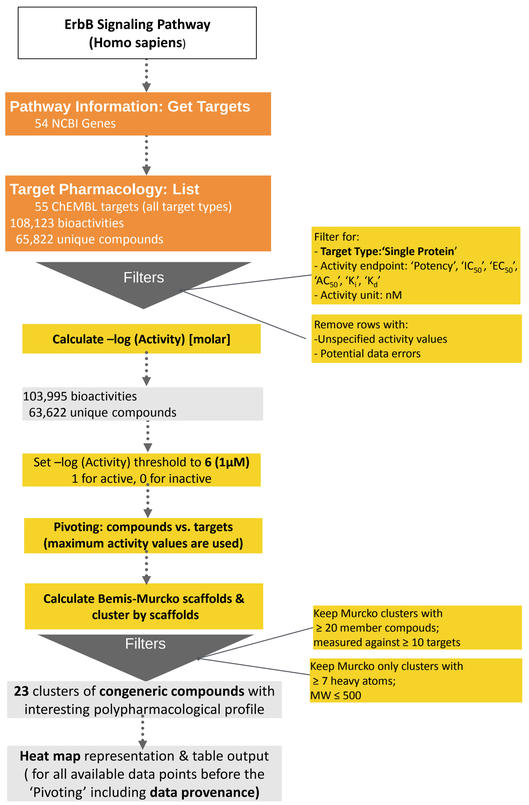
Background

- Finding new drug candidates that have the potential to become effective medicines has become increasingly challenging as our understanding of the relevant disease biology has developed.
- Whereas in the past the 'Holy Grail' was to identify molecules that selectively modulate just one biological target, there is now the realisation that complex diseases require more advanced approaches, for example, investigations that include the extension of traditional single-target structure-activity-relationship (SAR) studies towards delineating approaches, which deliver therapeutically relevant multi-target activity profiles. This strategy offers new opportunities towards a more holistic and mechanistic view of target modulation by small molecules in biological context.
- Since open life science data and collaborative drug discovery projects are enabling academic researchers to enter a world of increasing abilities when mining, interpreting, and extrapolating data, concurrently a need for computational approaches that help to manage and process the vast amount of information is emerging.
- In our previous study we delivered a workflow for identification of targets and compounds in pathways that were related to certain diseases by enriching the data with Gene Ontology (GO) and ChEBI annotations [1].

Aim of the study

- To obtain a convenient, fully flexible workflow for visualizing and retrieving multi-target SAR datasets in an automated fashion, which is able to deliver small molecule information in the context of biological pathways.

Workflow: using Open PHACTS API calls



Scheme of the applied workflow: *OPS version 1.3 (ChEMBL_16) and KNIME 2.9.4 was used*

- Retrieving small compound bioactivity data from Open PHACTS Discovery Platform [2] for all ‚Single Protein' targets in the ErbB Signalling pathway by using Open PHACTS API calls embedded in a KNIME workflow [3]
- Clustering of the data on basis of Bemis-Mucko framework [4] identity
- Heat map representation
- Inspecting SAR profiles of congeneric data sets

Research questions

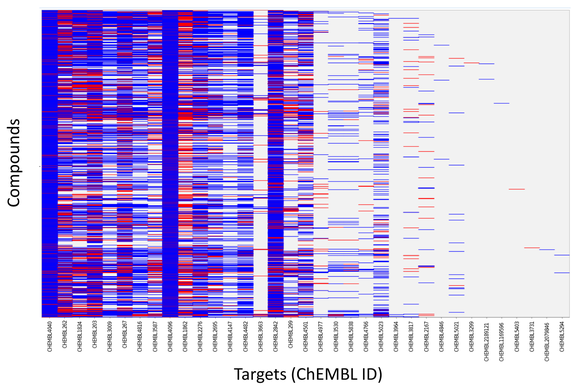
Does clustering by Bemis-Murcko frameworks lead to series of compounds.....

1. .....tested against the same targets?
2. .....originating from various publications?

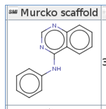
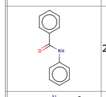
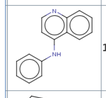
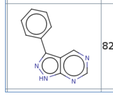
How do (poly)pharmacological patterns look like? → SAR studies

➢ Open PHACTS KNIME nodes can be retrieved from: http://github.com/openphacts/OPS-Knime

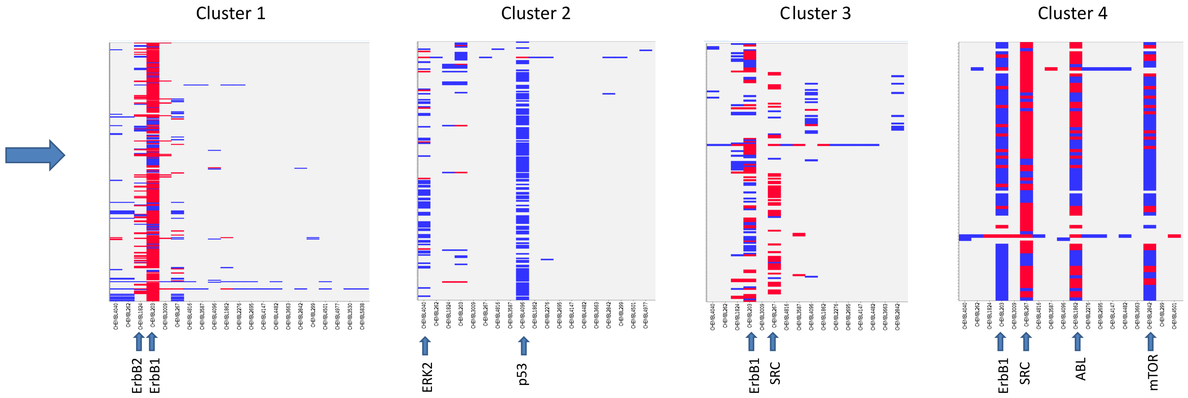
Results & Discussion

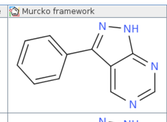
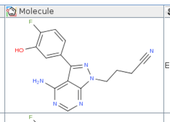
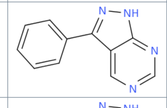
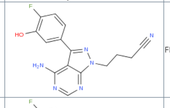
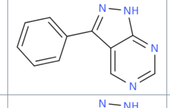
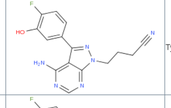
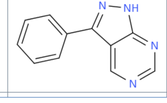
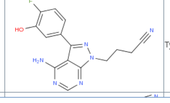


Full heatmap representation *of ‚single protein' targets vs compounds from ErbB signalling pathway (human) after ‚Pivoting' in binary format (red: active; blue: inactive; grey: no measurement): OPS version 1.3 (ChEMBL_16) and KNIME 2.9.4 was used*

cluster	Murcko scaffold	Compounds	Targets measured	Targets active	bioactivities
1		388	19	7	1,031
2		220	12	4	325
3		169	14	6	351
4		82	16	8	325

Four predominant Murcko frameworks in ErbB signalling pathway: *according to the number of available compounds (including the number of targets and number of active targets)*



CHEMBL Molecule	Murcko framework	Molecule	Target Name	CHEMBL Target ID	Assay CHEMBL ID	Act.Type	Relation	Value	Unit	PubMed ID	Act.Label
CHEMBL1242469			Epidermal growth factor receptor erb...	CHEMBL203	CHEMBL1243977	IC50	=	18,000	nM	18849971	0
CHEMBL1242469			FK506 binding protein 12	CHEMBL2842	CHEMBL1243969	IC50	=	2,200	nM	18849971	0
CHEMBL1242469			Tyrosine-protein kinase ABL	CHEMBL1862	CHEMBL1243972	IC50	=	963	nM	18849971	1
CHEMBL1242469			Tyrosine-protein kinase SRC	CHEMBL267	CHEMBL1243974	IC50	=	90	nM	18849971	1

Cluster 4 full table output: *For each compound in the cluster, all primary bioactivity entries are retained, including information about the activity type, the exact activity value, assay conditions, and provenance of the data (PMID's or DOI's)!*

Data provenance

- for some clusters (e.g. Cluster 4), the majority of the bioactivities originate from one paper: e.g. PMID 18849971 in Cluster 4
- other clusters (e.g. Cluster 1) are composed of bioactivity entries from a great variety a different publications:

e.g. 93 different PMID's for just ErbB1 in Cluster 1

Conclusions

- Integrating small compound data from whole pathways of interest delivers a complex picture of polypharmacological profiles in their biological context.
- Clustering these data by Bemis-Murcko scaffolds is a useful way to detect congeneric SAR series of compounds.
- Since integration of data from different providers is an essential element of Open PHACTS, **data provenance is of huge importance**.

As a consequence, our workflow not only
➢ delivers congeneric datasets to be used for (multi-label) SAR studies,
➢ but it is also capable of conveniently comparing assay descriptions, or underlying literature sources.