Compound Library Preparation in Galaxy: Application to HT Docking

Lucas X, Grüning BA, Senger C, Günther S

Pharmaceutical Bioinformatics, Institute of Pharmaceutical Sciences,
Albert-Ludwigs-University, Hermann-Herder-Str. 9, 79104 Freiburg i.Br., Germany
e-mail: xavier.lucas@pharmazie.uni-freiburg.de

Introduction: A variety of software tools and components exists for compounds analysis and drug discovery research. This includes tools for ligand- and structure-based *in-silico* screenings. In high-throughput (HT) docking large amounts of compounds are tested virtually and their activity is assessed, with the aim to distinguish between active and inactive molecules. The selection of the compounds to be included in the screening step is crucial. On one hand, large libraries have higher computational requirements. On the other hand, with small focused libraries one may omit interesting compounds belonging to novel classes. As HT docking methods become more common in the pharmaceutical research community, the preparation of custom libraries of compounds demands fully traceable workflows that ensure the repeatability of the results.

Thus, several tools for library preparation were integrated into a well-established workflow management system. Subsequently, these tools were applied in a HT docking experiment aiming to evaluate the performance of our HT docking protocol in the detection of known inhibitors of BET bromodomain-containing proteins out of large sets of compounds.

Methods & Results: Several tools for library preparation and compound selection were implemented in the Galaxy workflow management system (http://galaxy.psu.edu) as part of a cheminformatics package to enhance and facilitate

- 2D similarity searches of compounds,
- substructure searches,
- retrieval of physico-chemical properties for specific compounds, and
- filtering of large databases of small comounds by applying user-defined and commonly used sets of physico-chemical constraints (e.g. Lipinski's Rule of Five or Lead-Like properties [1]).

On the basis of the ZINC (http://zinc.docking.org) [2] and ChEMBL (https://www.ebi.ac.uk/chembldb) [3] compound libraries the implemented tools were used to test the sensitivity of the HT docking protocol. The ability to identify known

binders of BET bromodomains [4,5] with Glide (Schrödinger Inc.) was tested with different parameters and yielded in interesting results. Depending on the library design and Glide parameter setting most known binders were ranked onto the top of all docked compounds. Highly ranked compounds were identified as potential novel inhibitors of the specific target. Calculations were carried out in the Baden-Württemberg Grid, in Germany. Using our docking protocol, large libraries of compounds can be screened in few hours.

Availability and Future Prospects: The Galaxy cheminformatics package is available for third parties on request. In the near future further HT docking screening tools are planned to be implemented, e.g. compound clustering methods for large libraries or pharmacophore-based searches. Our HT docking protocol will be used for the identification of novel inhibitors of pharmaceutically relevant protein targets and protein-protein interactions.

References:
1. Teague SJ, Davis AM *et al* (1999). Angew. Chem. Ind. Ed. Engl. 38(24):3743-48.
2. Irwin JJ and Shoichet BK (2005). J. Chem. Inf. Model. 45(1):177-82.
3. Wendy A (2009). J Comput Aided Mol Des. 23(4):195-8.
4. Filippakoulos P, Qi J, Picaud S *et al*. (2010). Nature. 468(7327): 1067-73.
5. Chung C, Coste H, White JH *et al*. (2011). J. Med. Chem. 54: 3827-38.