Galaxy – a tool for automation and data management in computational drug discovery

As chemistry enters the digital era, the development of computational tools and infrastructure to support scientists becomes more and more critical. Within the NFDI4Chem initiative, this digital transformation is being accelerated by the adoption of advanced platforms such as Galaxy. This talk aims to demonstrate how Galaxy, a flexible, web-based environment for data analysis, supports the chemistry research community through access to cheminformatics applications and integration with research data infrastructure.

At Gain Therapeutics, Galaxy serves as a workflow management tool for Magellan™, our computational drug discovery platform. Magellan™ comprises a wide variety of physics-based and AI tools, from target assessment to virtual screening and compound rescoring; these tools possess diverse computational requirements, including GPU hardware and remote cluster execution. Galaxy enables these tools to be seamlessly combined into a single workflow, allowing virtual screening of targets to be launched at the press of a button. This level of automation facilitates the screening of ultra-large chemical spaces that have become available in recent years, such as Enamine REAL, containing up to trillions of compounds.

In addition, Galaxy provides an intuitive, user-friendly way to publish open-source scientific tools. For example, LiraSearch, a tool for three-dimensional similarity search of the ZINC database developed at Gain, is now available as an open-source application on the European Galaxy server. The talk will explain the LiraSearch algorithm and its potential use-cases for drug discovery scientists.

Finally, we will discuss the practical application of FAIR data principles (Findable, Accessible, Interoperable and Re-usable) using RO-Crates, EDAM ontologies, and metadata standards. These approaches enable sharing and reusability of chemical data and analysis pipelines. The talk will discuss how Galaxy can interface with research data infrastructures, supporting federated services in the NFDI ecosystem and giving chemists the ability to build and run pipelines covering the whole of the data analysis process, from data generation to archiving and reuse. These developments allow chemists to make use of cutting-edge computational resources and open data, promoting collaboration and innovation.