

LSM-Webinar

Time: THU April 8, 2021 02:00 PM Amsterdam, Berlin, Rome, Stockholm, Vienna

Join Zoom Meeting

<https://psich.zoom.us/j/96912594075>

New data tools for biological research at PSI

Spencer Bliven

LSM/BIO

The daily practice of science requires the completion of numerous tasks for data management and analysis. Automating and streamlining such tasks can have a large effect when used by many users. Here I present a number of tools developed in collaboration with scientists in BIO and LSM to streamline workflows, both at PSI and externally.

1. **RCSBsearch** (<https://rcsbsearch.readthedocs.io>) is a python package to facilitate identification of protein structures in the RCSB Protein Data Bank based on an extensive set of attributes, including structural metadata, ligand and drug binding, structural similarity, and sequence motifs. The package was applied to the identification of potential protein scaffolds derived from viral particles based on oligomeric state, protein size, quaternary symmetry, and other considerations. Protein designs based on the results could be useful for antigen exposure in vaccine development, protein binding assays, and structural biology.

2. The Data Catalog is the PSI-mandated repository for archiving experimental and scientific data. However, preparing data for ingestion into the catalog can be difficult for users that are not familiar with the JSON format and recommended ontologies. Two utilities have been developed to facilitate creation of the metadata.json file required to archive new datasets. The **SciCat Metadata Editor** (https://bliven_s.gitpages.psi.ch/SciCatEditor) provides a web-based editor for the metadata files. It provides templates for common data types, and validates metadata files against the schema before downloading. The **mxarchive** tool (https://git.psi.ch/bliven_s/mxarchive) can directly generate metadata files for macromolecular crystallography projects based on the standard directory layout.

3. PSI has significant expertise regarding G-protein coupled receptors (GPCR). An ongoing research project hopes aims to understand GPCR binding to G-proteins based on structural comparison and deep learning. As part of the feature extraction, we have developed a python package (<https://github.com/sbliven/gpcrdb>) to interface with the **GPCRdb**, which provides GPCR annotations, alignments, binding affinities, and other data.

The tools are developed closely with scientists to fulfill immediate needs. However, they are also made available open source whenever possible to encourage use by the broader community.

<https://indico.psi.ch/event/11197/>

Organized by:

Laboratory for Scientific Computing and Modelling

Dr. Derek Feichtinger

derek.feichtinger@psi.ch