

## Building a FAIR Repository for Fragment-Based Screening Campaigns

G. L. Evans<sup>1</sup>, A. Tanweer<sup>1</sup>, J. Tolchard<sup>2</sup>, D. Harrus<sup>1</sup>, P. Choudhary<sup>1</sup>, J. Flemming<sup>1</sup>, S. Velankar<sup>1</sup>

<sup>1</sup>Protein Data Bank in Europe, European Molecular Biology Laboratory, European Bioinformatics Institute (EMBL-EBI), Wellcome Genome Campus, Hinxton, Cambridge CB10 1SD, UK, <sup>2</sup>Center of Nuclear Magnetic Resonance at Very High Fields (Centre de Résonance Magnétique Nucléaire à Très Hauts Champs - CRMN), 5 Rue de la Doua, 69100 Villeurbanne, France

Email: [gle@ebi.ac.uk](mailto:gle@ebi.ac.uk) or [pdbehelp@ebi.ac.uk](mailto:pdbehelp@ebi.ac.uk)

Fragment-Based Screening (FBS) is a drug discovery technique using small chemicals with low affinity when screened against biological targets. FBS is more cost-effective than traditional High-Throughput Screening (HTS), but archiving FBS results is a multi-faceted challenge, with only a few positive results deposited to public databases.

For over 50 years, the Protein Data Bank (PDB) has served as the global archive for experimentally-determined 3D biomacromolecular structures. Managed by the Worldwide PDB (wwPDB) consortium, the PDB provides a vital resource for structural biology and drug discovery, offering open access to high-quality structural data via unique identifiers, standard formats, and various retrieval methods including REST-API for programmatic access and web-based portals. FBS can be performed with various experimental methods, including 3D biomacromolecular structures at atomic resolution. Technological advances have enabled FBS campaigns to generate hundreds of protein structures, with most results, particularly negative data, not deposited to public databases.

To address this gap, we are building a FAIR (Findable, Accessible, Interoperable, and Reusable) repository to capture positive and negative results and the chemical library screened and thus provide a structured framework to identify and connect data from the same FBS experiment. Leveraging the wwPDB's data standard framework, this repository promotes transparency, reproducibility, and data reuse. An application called mmcif-gen, indexed by PyPi (<https://pypi.org/project/mmcif-gen/>), has been developed to generate information in the wwPDB's data standard framework (schema: <https://github.com/PDBEurope/InvestigationCIF>).

A pilot repository is available at: [https://ftp.ebi.ac.uk/pub/databases/msd/fragment\\_screening/investigations](https://ftp.ebi.ac.uk/pub/databases/msd/fragment_screening/investigations)