OSCARS
Open Science Clusters' Action
for Research & Society

Funded Project

Implementing FAIRness in structure-based drug design through Fragalysis Cloud

Presenter: Warren Thompson, Diamond Light Source, https://orcid.org/0000-0003-1474-7810

Implemented by

diamond

XCHEM
Diamond
Fragment
Screening

INFORMATICS
MATTERS
SOLUTIONS FOR DRUG DISCOVERY

UKRI
Research Complex
at Harwell

M2M
SOLUTIONS

Funded by
the European Union

CHALLENGE ADDRESSED

1. Haphazard data dissemination

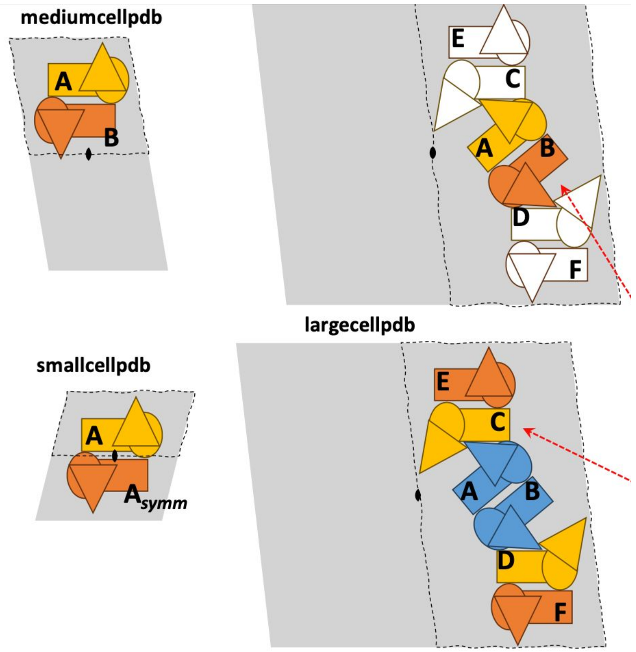
- Structural and activity data remain scattered, uncurated, and difficult to deposit at scale.
- Technical bottlenecks prevent users from delivering large-volume, FAIR-ready datasets to PDBe/ChEMBL.

2. Lack of integrated tools for collaborative SBDD

- Iterative DMTA workflows rely on fragmented, often proprietary tools.
- Computational methods are powerful but inaccessible, hard to deploy, and not linked to experimental data.

1. Developed new tools with in-house expertise

- Leveraged XChem's structural biology and crystallography expertise to define biological assemblies, binding sites, and curation rules for uploading data to Fragalysis



2. Agile, feedback-driven development

- Ran continuous feedback loops between developers, crystallographers, and medicinal chemists.

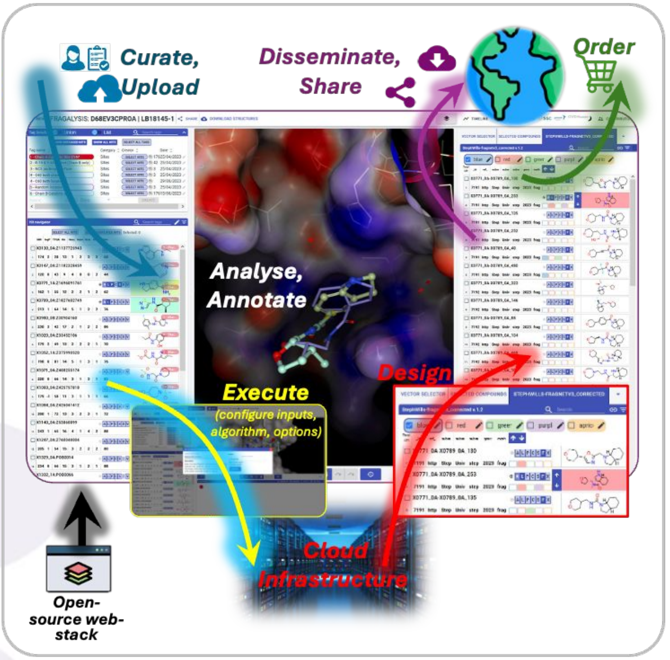
3. Robust data-curation workflows

- Developed XChem Align to collate and interpret crystallographic data in biological reference frames.

RESULTS

Task 1 – FAIR Structural Data

- XChemAlign generated aligned, biologically meaningful datasets.
- Thousands of structures released for **nine ASAP antiviral targets**.
- Automated deposition workflows via **OpenBind** & PDBe.



Task 2 – FAIR Programmatic Access

- REST APIs enable automated data retrieval and upload.
- Standardised routes for sharing structural and activity data.
- Supports community analysis, integration, and reuse.

Task 3 – Compute Workflows

- Squonk notebooks provide installation-free access to algorithms.
- Fragment progression algorithms directly connected to Fragalysis data.
- Enables rapid design cycles and AI/ML-ready datasets.

RESULTS

OSCARS

Sustainability

- **FAIR-by-design infrastructure**: All structural & activity data stored with versioning, provenance, and programmatic access via REST & Python APIs.
- **Standardised datasets**: XChemAlign ensures aligned, biologically meaningful reference frames reusable across projects and years.
- **Integrated deposition workflows**: Collaboration with PDBe via **OpenBind** embeds our pipelines into community standards.

Community exploitation

- **Open computational ecosystem:** Squonk notebooks give algorithm developers a platform to deploy their fragment progression without User installation barriers.
- **Reusable curated datasets:** Aligned binding sites & metadata support AI training, benchmarking, and meta-analysis across targets.
- **Shared views & snapshots:** URL-based structural views allow precise, reproducible collaboration across distributed teams.

TEAM

OSCARS

XCHEM Diamond Fragment Screening

INFORMATICS MATTERS
SOLUTIONS FOR DRUG DISCOVERY

M2M SOLUTIONS

PDBe
Protein Data Bank in Europe

Frank von Delft

Tim Dudgeon

Boris Kovar

Ahsan Tanweer

Warren Thompson

Alan Christie

Matej Vavrek

Genevieve Evans

Max Winokan