Global Phasing's viewpoint on the ISPyB collaboration and its evolution

Gerard Bricogne, Rasmus Fogh, Peter Keller, Clemens Vonrhein

We are not a synchrotron but we have an extensive area of contact with synchrotrons and their end-users

- We are supported by an international **Consortium of drug discovery companies** (Big Pharma, and also CROs doing outsourced work for them – 22 members in total).
- We are **developers of** application programs **autoPROC** and **STARANISO** (for diffraction image processing), **BUSTER** (for structure refinement) and **Pipedream** (designed for use in highly integrated environments and high-throughput **operations that interact strongly with synchrotrons**, e.g. corporate pipelines and CRIMS at EMBL-HTX).
- We have developed **workflows** combining a high degree of **automation** with a high level of **built-in expertise** for designing **optimised data collection experiments** for both experimental phasing data and native data and **directly controlling their execution**.
- We are taking part in the development of **capabilities to deposit** into the PDB , or into corporate databases, not just model coordinates but also the associated **diffraction data, accompanied by the most complete metadata and quality metrics possible**.

- **In making available our application program autoPROC for use at synchrotron auto-processing facilities**, we have had to assist in the specification of each beamline configuration input into the program, and in the fullest presentation of processing results to users. This has brought us into close contact with developers of these auto-processing systems and of both ISPyB and of its frontends.
- We participated in the HDRMX exercise of codifying a metadata “gold standard” that can guarantee **transferability of reprocessing**, which is crucial to making archived raw data actually re-useable in the future.
- The development of STARANISO (with now nearly 10,000 successful submissions to our Web server, and also included in autoPROC) has led us to produce **new data characteristics and metrics that break the mould of Table 1, of ISPyB and of its frontends**, requiring an extension of tables and of display layouts (also of the PDBx/mmCIF dictionary).

- **In the development of expert automated workflows**, we have had to cope with the diversity of beamline and end-station components and of their control software. This has brought us into a close interaction with the MXCuBE developers' community, including taking part in the sharing of housekeeping tasks and in the large ongoing code refactoring effort.
- Our work in this area has shown the feasibility of achieving **transferability of expert automation** through an emphasis on abstraction, enabling our workflows for translational calibration, diffractometer calibration and data collection for both native and experimental phasing to run variously on several beamlines (1 under GDA, 4 under MXCuBE v2).
- However **the presentation of the processing results** produced by autoPROC **for such experiments** (with several scans in different orientations, possibly at several wavelengths and with certain patterns of interleaving) **is still beyond the capabilities of ISPyB and its frontends**.

- **In our contribution to the creation of a highly integrated HT screening resource with the EMBL-HTX**, designed to use several synchrotrons simultaneously in a single project, and also in our general interaction with our Consortium members, we have become acutely aware of **how crucial ISPyB is in operating such a facility** and of how highly desirable it would be to be able to ship samples, request data collection and access the data themselves and associated processing results in **a uniform, consistent manner across multiple synchrotrons**.
- Achieving this would be important in the perspective of the staged synchrotron shutdowns in the future, to facilitate the creation of a genuine **synchrotron grid** to ensure minimum disruption and loss of capacity for industrial users when such shutdowns occur.

- **In our developments to enable the archiving of diffraction data** along with the atomic models that have been refined against them, we are working with the PDBe and the wwPDB on how to extend the PDBx/mmCIF dictionary to accommodate definitions of **new data quality metrics** (e.g. from STARANISO) and **new metadata** for composite datasets so that they can be archived in the PDB.
- These **should therefore be archived in the first place** as dataset attributes and as auto-processing results **by ISPyB itself**, and made user-accessible through its frontends to enable submission to the PDB or deposition into internal corporate databases.
- This provides yet another source of **demand for a standardisation of ISPyB's frontends and for consistency in their evolution**.

Conclusion

- We have a big stake in the achievement of **standardisation and transferability** of know-how and capabilities **across synchrotrons**.
- By virtue of our position, **we are especially sensitive to heterogeneity between synchrotrons**, especially regarding MXCuBE, auto-processing and ISPyB.
- In our view **the continuation of the ISPyB collaboration** and an evolution towards greater versatility and interoperability of its MX frontends **is absolutely vital** to the future strength of SB and SBDD in Europe.
- **We can contribute resources** to improving the status quo wherever this can be expected to have a beneficial impact on the operations of our Consortium members – and ISPyB certainly fits that description.