

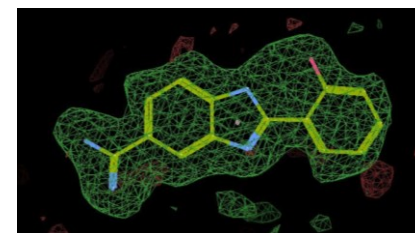
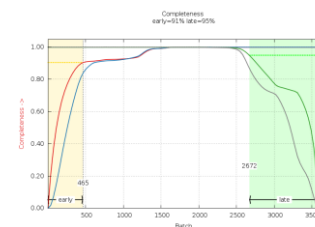
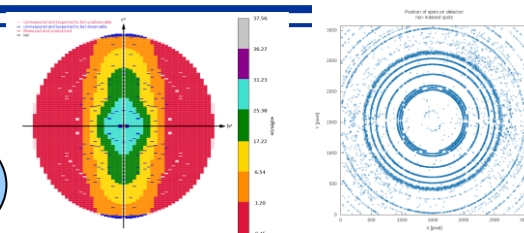
Status Report for Global Phasing Ltd.

Gerard Bricogne & Collaborators

Global Phasing Ltd, Cambridge, UK

What we do

- ❑ Data collection and experiment
 - Global Phasing **workflow** (MXCuBE, GDA, ISPyB)
 - **Multi-sweep strategies** for native and phasing data collection
- ❑ Data processing
 - **autoPROC** (XDS, CCP4, with much “added value”)
 - **STARANISO** (anisotropy characterisation and mitigation)
 - multi-lattice, ice rings, pruning poor image ranges, bad pixels
- ❑ Ligands
 - **BUSTER** (ligand detection maps)
 - **Grade** restraints dictionary generation (CSD, QM)
 - **RhoFit** (automated fitting into density)
- ❑ Refinement
 - **BUSTER** (structure refinement)
 - **Pipedream** (ligand detection & refinement pipeline)
- ❑ Validation, reporting and deposition
 - **buster-report**, **CRIMS-Pipedream interface**
 - **PDBx/mmCIF** output combining model and rich data
 - Advanced methods for data validation and “auditing”



Figure

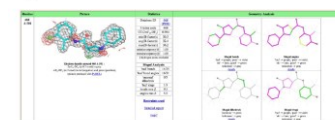


Figure 1. Example highlighting the value of presenting ligand electron density model fit and geometrical analysis from CCDC Mogul from the Global Phasing Buster Report (PDB ID: 2H7P, later superseded by entry 4TZT (He et al., 2006); CCD ID: 488).

Our interest in ISPyB: overview

- We are not a synchrotron, we are application developers, the relevant ones in the ISPyB context being the data collection Workflow, autoPROC, and STARANISO.
 - We do have users, though, but very many of them only see autoPROC results as produced by the synchrotron auto-processing pipelines, archived in the ISPyB backend, and presented by its frontend(s).
 - ISPyB therefore plays a central role in our ability to deliver the best value to our users, both academic and industrial, in a way that is inextricably linked to auto-processing facilities.
 - It is important for us that autoPROC be run according to best practice, and that its treatment at synchrotrons who develop their own processing software be equitable (which is not always the case).
-

Our interest in ISPyB: a closer look

- Through the development and deployment of our Workflow, we are also generators of data collection protocols whose execution creates several datasets inter-related through high-level metadata
 - These inter-relations cannot be represented in ISPyB, and the processing by autoPROC of such composite datasets produces a set of results for which no adequate presentation exists so far in the ISPyB frontends.
 - there have always been difficulties accommodating e.g. multi-sweep datasets in ISPyB front-ends. Currently the results for individual sweeps can be displayed, except for the last one because its display slot is used for the results of the combined sweeps.
 - this is shoe-horning rather than accommodation through proper expansion of functionalities.
 - The outdated “*one sample -> one dataset -> one processing result*” paradigm still rules
 - The inclusion of autoprocessing within the scope of the ISPyB collaboration seems a long overdue item (a viewpoint we put forward at both of the 2018 meetings).
-

Previous way of looking to the future

- Two circumstances seemed to provide an opportunity for change:
 - the development of Py-ISPyB
 - the need to support data archival and display of processing results for e.g. serial crystallography and other data collection techniques involving multiple inter-related datasets
 - We volunteered to contribute scientific input into the design of the MX part of Py-ISPyB, but our availability towards this activity would have involved a significant induction period, and the Py-ISPyB development activities started with authentication and SSX, then (from our external standpoint) went silent at the end of 2022.
 - A significant call on our resources at the time was the not-unrelated task of contributing to the design of a set of extensions to the mmCIF dictionary to archive unmerged data capable of accommodating both data collected by the rotation method and serial crystallography data.
 - we prepared the ground for Py-ISPyB developers and the wwPDB SubGroup on Data Collection and Processing to meet in a VC, but this didn't happen yet
-

New way of looking to the future

- The outcome of ESRF's Pilot Project of rebasing of ISPyB on ICAT, announced at the SOLEIL meeting, has been keenly awaited.
 - As our work with the group of Ashwin Chari (MPG, Goettingen) at the EMBL-HH P14 beamline (Gleb Bourenkov) using the GPhL Workflow has produced an abundance of multi-sweep datasets and associated autoPROC results that ISPyB was unable to properly accommodate, we (Rasmus, with Gleb & Ashwin) produced a hand-crafted facility to do this (Rasmus's talk).
 - We are open-minded about the necessary metamorphosis of ISPyB.
 - A related advance in conceptualising the requirements for such a metamorphosis has been the idea of an "Abstract LIMS" proposed by the MXCuBE Developers, as this offers the possibility of preserving the natural structure of the data and metadata produced in the course of an experiment, e.g. as driven by the GPhL Workflow (including high-level strategy information) and of the derived processing results.
 - We also need to bear in mind the need to present as uniform as possible an external appearance to users whose projects involve multiple synchrotrons.
-