Achieving higher performance in high-throughput compound and fragment screening campaigns by the use of "Club Class" data collection with Pipedream and CRIMS

Gérard Bricogne[1], Leigh Carter[1], Claus Flensburg[1], Rasmus Fogh[1], Peter Keller[1], Wlodek Paciorek[1], Andrew Sharff[1], Clemens Vonrhein[1], Irina Cornaciu[2], Michal Jamroz[2], Jose A. Marquez[2], Martin Lehmann[3], Djordje Musil[3]

[1] Global Phasing Limited., Sheraton House, Castle Park, Cambridge CB3 0AX, United Kingdom,
[2] EMBL Grenoble Outstation, 71, Avenue des Martyrs, 38000 Grenoble, France,
[3] Discovery Technologies, Merck KGaA, Frankfurter Str. 250, D-64293 Darmstadt, Germany.

Fourth-generation X-ray sources, fast sample delivery systems for micro-crystals and pixel-array detectors with ever faster frame rates are radically extending ways of collecting diffraction data, while associated software developments are creating a continuum of options between serial experiments at synchrotrons and at XFELs.
From the viewpoint of structure-based drug discovery, these developments will give better access to targets for which "macro-crystals" cannot be obtained. However, they will do so at a greatly increased cost, both in terms of beam time and computational resources and of the refactoring of sample production workflows. This encourages renewed efforts to more fully exploit the intrinsic advantages of macro-crystals, when available, through better experiments. Unfortunately, exclusive emphasis on automation and speed in high-throughput data collection workflows has left little room for invoking any form of expert decision making in data collection without considerable disruption of automation.
An unlikely opportunity linked to the development of the Diamond I23 experimental phasing beamline (Armin Wagner) has enabled us to overcome this limitation via a "Third-party design and control" approach that makes our solution transferable between different beamline software environments with minimum effort. This approach has been validated by the live execution of our workflows on Diamond beamlines running under the GDA control software, and work is ongoing to transfer them to the MXCuBE control software in use at all European synchrotrons.
We have been able to demonstrate that "Club Class" multi-orientation datasets can significantly enhance the detection and characterisation of small weakly bound fragments - a process fundamental to fragment-based drug discovery - in the context of a collaboration with the EMBL and Merck-Serono teams in which the capabilities of the HTX facility at the EMBL Grenoble outstation and of the CrystalDirect technology were applied to a fragment screening campaign. Datasets collected on the ESRF MASSIF-1 beamline were processed with the Global Phasing "Pipedream" software suite that seamlessly combines automated data reduction and model refinement with ligand density elicitation and automated ligand fitting. New interfaces for the ranking and evaluation of Pipedream results were developed within the Crystallization Information Management System (CRIMS) in order to support rapid analysis and decision making in the context of large-scale compound and fragment screening campaigns. The use of a small number of “Club Class” datasets (manually collected on the ESRF ID30B and SLS PX-III beamlines) led to a significant increase in the number of hits and enabled a complete elucidation of initially ambiguous binding modes for a number of challenging fragments. The approach described here is widely applicable and illustrates the potential of the combined use of “Club Class” data collection strategies and of fully automated protein-to-structure X-ray crystallography pipelines to achieve higher performance in large-scale compound and fragment screening campaigns.