

MXCuBE-related developments at GPhL:
their basis in GDA-related *transferable*
developments

Gerard Bricogne,
Claus Flensburg, Wlodek Paciorek,
Peter Keller, Rasmus Fogh,
Clemens Vonrhein

Global Phasing Ltd., Cambridge, UK

An underexploited niche:

“Club Class” data collection on macro-crystals

- Growing bigger crystals saves experimental and computational resources!
- What about trying to **make better use of “Humpty Dumpty in one piece”**? I.e. collect better data on conventional crystals by better exploiting the available technologies?
- The dominant trend has been in the other direction: the ever-increasing speed of MX beamline instrumentation has led to ever-stronger emphasis being placed on brevity of execution as the main design goal for data collection protocols, to the exclusion of other criteria that would aim at achieving higher data quality. This can be counter-productive. E.g. it has excluded multi-axis goniometers.
- Global Phasing, among others, has been interested (since the BIOXHIT project, 2004-2009) in bucking that trend by creating combined capabilities for the **fast design of optimal strategies and the direct supervision of their execution** on an actual beamline.
- Not convinced? You need to look closer at some data!

From anisotropy mitigation to data quality visualisation

The STARANISO Server

<http://staraniso.globalphasing.org/>

*Anisotropy of Diffraction Limits and
Bayesian Estimation of Structure Amplitudes*

GΦL

Global Phasing Limited

Main developer: Ian Tickle












Illicium verum

PERSPECTIVE

Against Method: Table 1—Cui Bono?

Bernhard Rupp  DOI: <https://doi.org/10.1016/j.str.2018.04.013> |  CrossMark[Article Info](#)

Switch to Standard View

-  PDF (652 KB)
-  Download Images (.ppt)
-  Email Article
-  Add to My Reading List
-  Export Citation
-  Create Citation Alert
-  Cited by in Scopus (0)
-  Request Permissions
-  Order Reprints (100 minimum order)

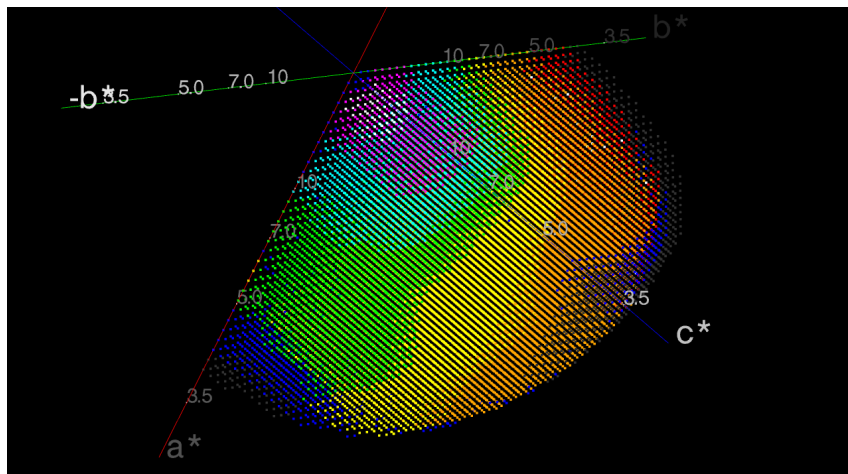
[Summary](#)[Full Text](#)[Images/Data](#)[References](#)[Related Articles](#)

The almost universally required “Table 1,” summarizing data-collection and data-processing statistics, has in its present form outlived its usefulness in almost all publications of biomolecular crystal structure reports. Information contained in “Table 1” is insufficient to evaluate or repeat the experiment; is redundant with information extractable from deposited diffraction data; and includes data items whose meaning is under increased scrutiny in the crystallographic community. Direct and consistent extraction and analysis of data quality metrics from preferably unmerged intensity data with graphical presentation of reciprocal space features, including impact on map and model features, should replace “Table 1.”

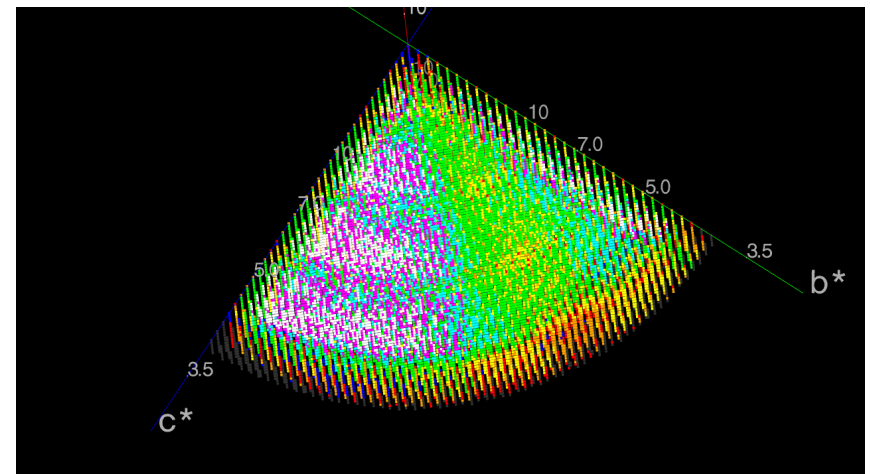
Keywords:

data collection statistics, diffraction data analysis, merging statistics, data quality indicators, reciprocal space, diffraction data presentation

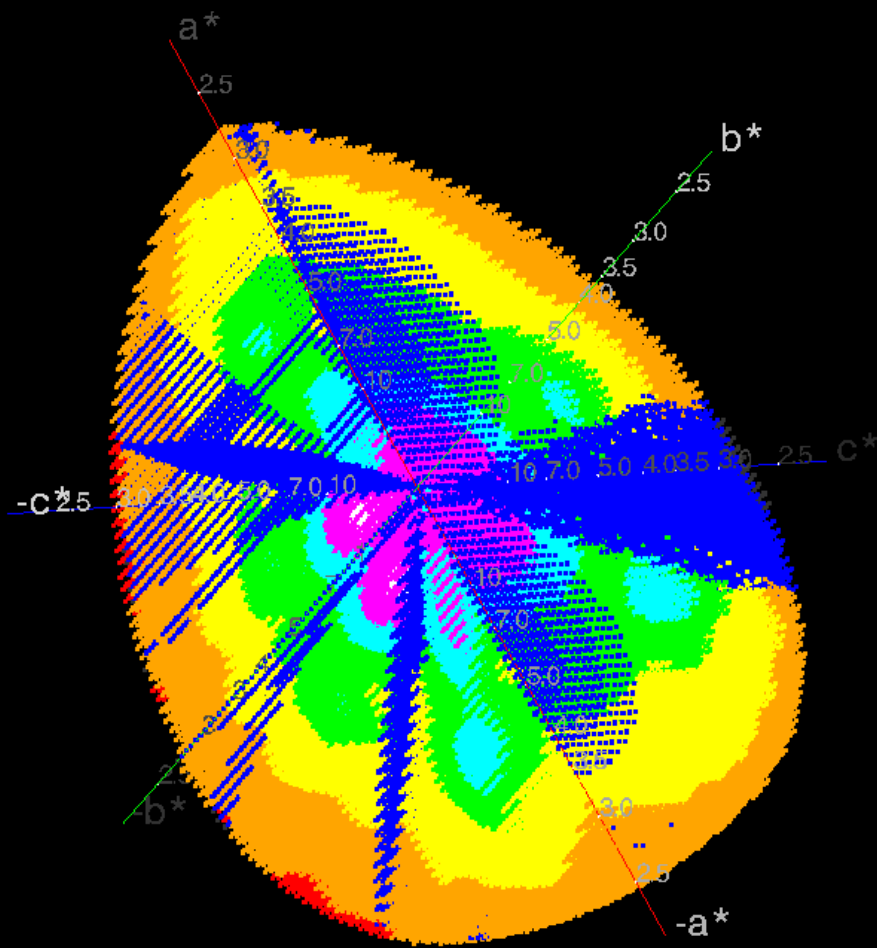
Local $\langle I/\sigma(I) \rangle$ and redundancy plots for Bernhard Rupp's example



Truncation of anisotropic data
by the edge of a detector set too far

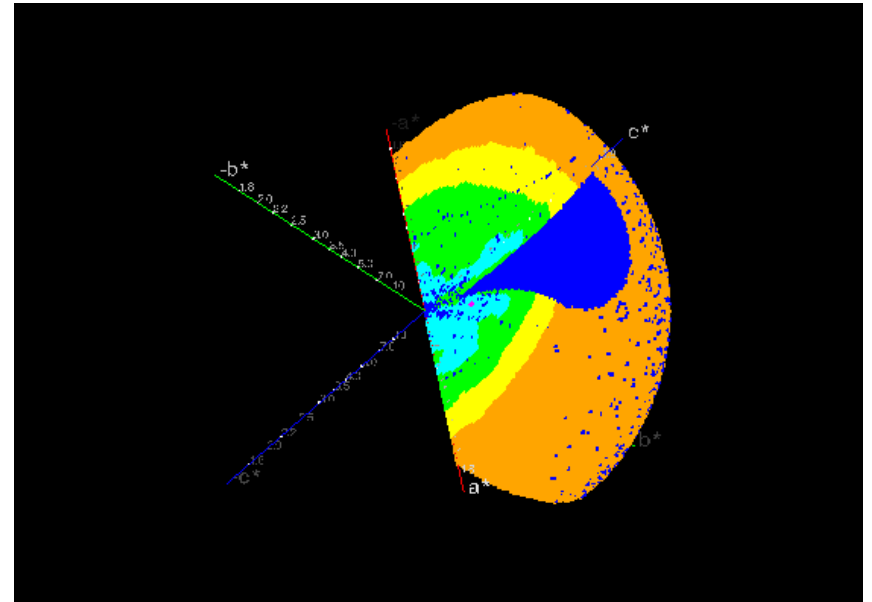
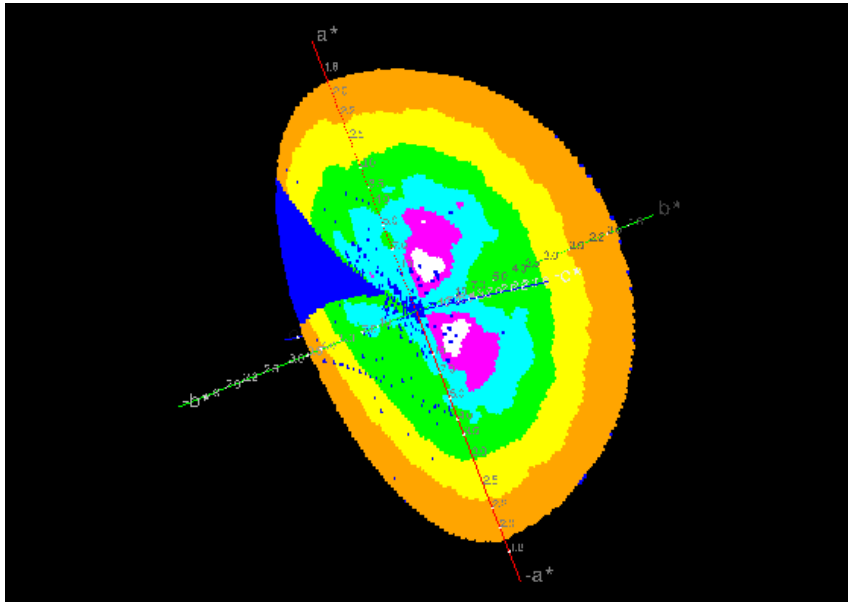


Non-uniform distribution of redundancy
(a.k.a. multiplicity) of measurement

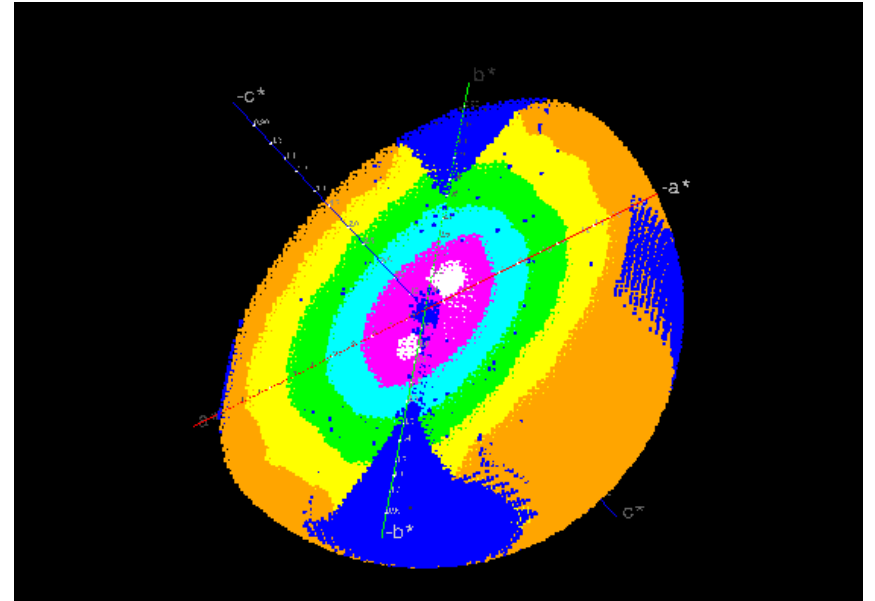
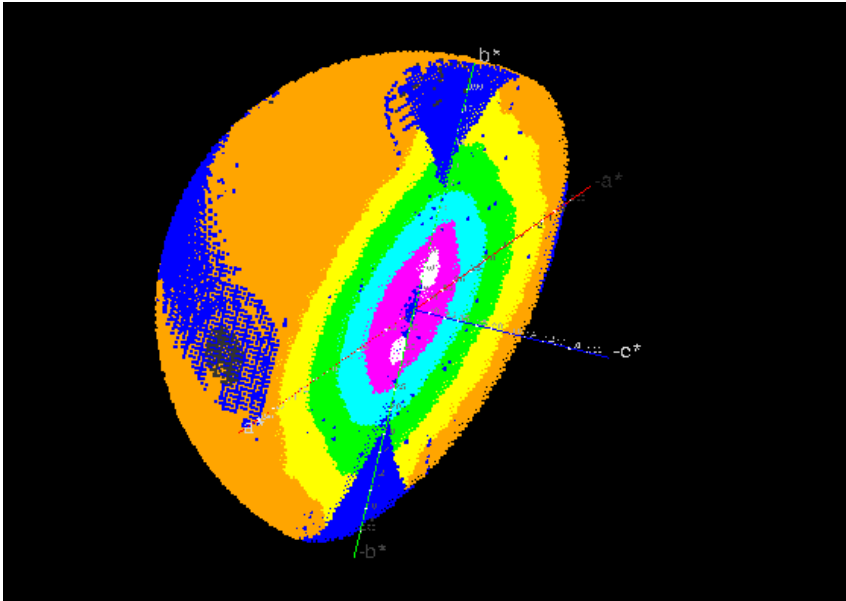


Shortcomings of experimental protocols
are repeatedly observed, e.g. *missing cusp*
and truncation by the detector edges

1YAR (20S proteasome data, 1.8A)



4GCA (aldose reductase, 0.9A)

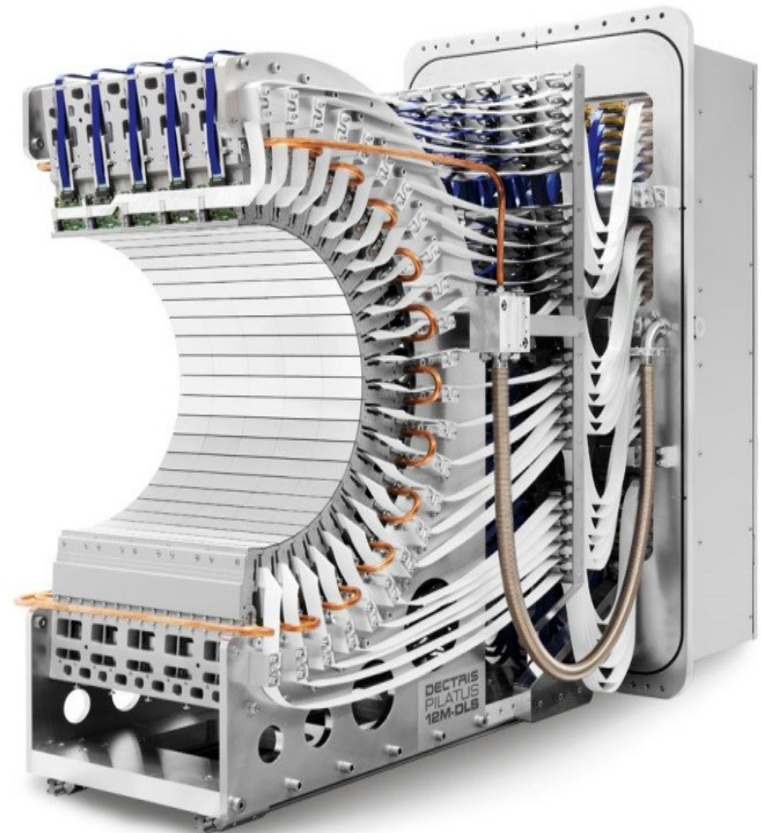


Can one improve experimental protocols *without time penalty nor loss of automation?*

- What expertise do we need to capture?
 - **Achieving completeness in spite of all impediments**
 - e.g “filling the cusp” for low-symmetry samples
 - requires the use of a **multi-axis goniometer** (mini-kappa, PRIGo, ...)
 - Achieving maximum data resolution
 - Eliminating systematic errors
 - e.g. taking advantage of partial cancellation of RD effects by using interleaved strategies and multi-axis goniometry
- Biggest obstacles:
 - **time and high-throughput pressures**, expediency
 - need for multiple improvements at consecutive steps before the benefits of any one of them can become visible
 - fragmentation of the necessary efforts by lack of uniformity across beamlines

An unlikely opportunity to “push the frontiers”: the Diamond I23 long-wavelength beamline (Armin Wagner)

- *In vacuo* operation
- Wavelengths 1.5-4.5Å and perhaps beyond (4.5Å has been achieved)
- Half-cylindrical Pilatus 12M detector
- Full kappa goniometer



The “Third-party Design and Control” paradigm

- Our proposed solution is to **capture expertise in transferably automated form** - to be pursued through
 - Instrument-generic applications (cf EEC Workshop) for simulation, prediction and strategy design
 - A Workflow capable of controlling the execution of a designed strategy by communicating with the Beamline Control Software in a generic manner thanks to an Abstract Beamline Interface
 - An off-site development methodology using Emulation
- **Roadmap:**
 - create a Workflow that can drive a live multi-wavelength experiment at Diamond under GDA and be transferred to MXCuBE.
 - Adapt to optimised native data collection protocols to produce **Club Class datasets**
- “Club Class” datasets with very high completeness and uniform distribution of $I/\sigma(I)$ should produce maps with lower levels of error and hence increase the sensitivity of detection of small, weakly bound fragments, especially for low-symmetry crystals.
- **This will contribute to more fully exploiting “macro-crystals”.**

30 Sep 2016: I04, Diamond, viewed from Sheraton House (GPhL)



The GPhL MX Experimental Workflow went on to execute an **interleaved 3-wavelength MAD experiment** on a lysozyme crystal whose four-fold symmetry axis had been aligned along the Omega axis using the mini-kappa of **the I04 beamline** via our strategy program StratCal.

Connecting to MXCuBE ...

First live contact: ESRF ID30B 10 Oct 2017



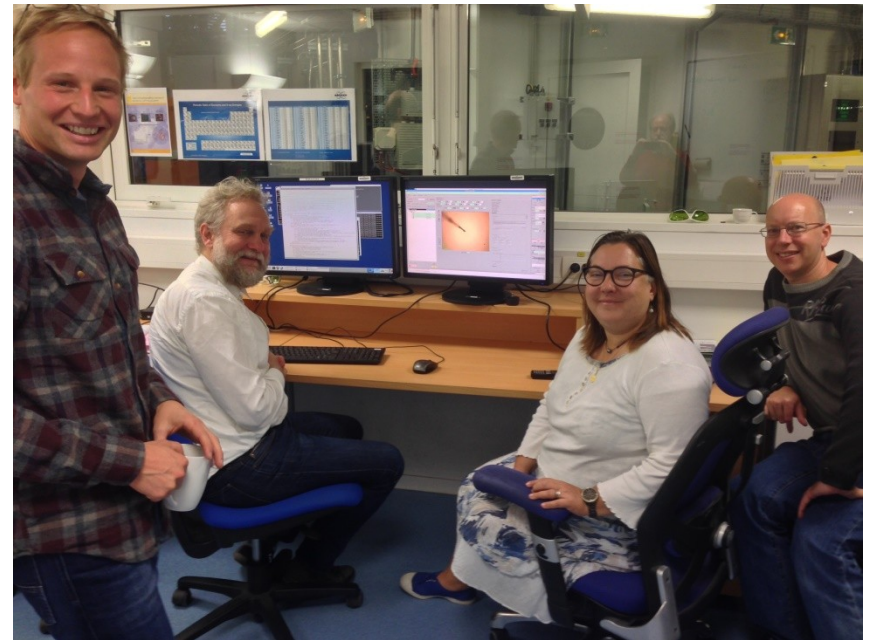
Translational Calibration on ID30B

12 Oct 2017

Working on it ...



Success!



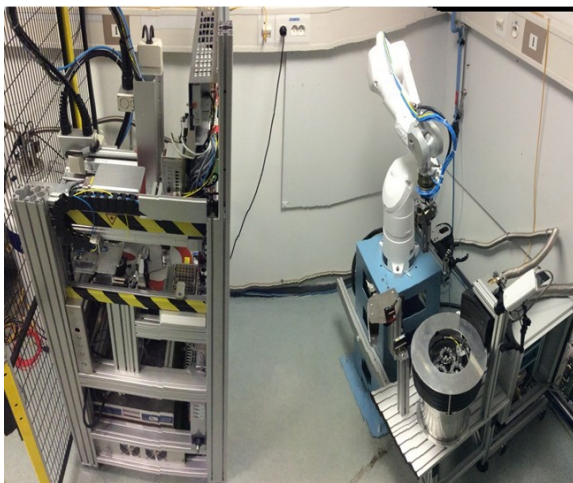
Running the experimental phasing workflow from GPhL on ID30B, 1 May 2018



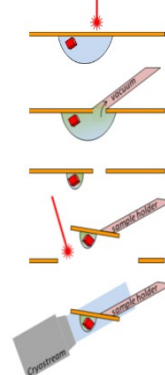
The EMBL-Grenoble Collaboration

- Initiated in 2016
- The direct connection between Jose Marquez's HTX Lab at EMBL and beamlines ID30B and MASSIF-1 at ESRF, using ISPyB for sample handling and data archiving, MXCuBE for data collection and Pipedream for data processing and ligand screening, offered a self-contained microcosm in which to test all our developments in one place.

CrystalDirect™ @ EMBL Grenoble
High throughput automated crystal processing



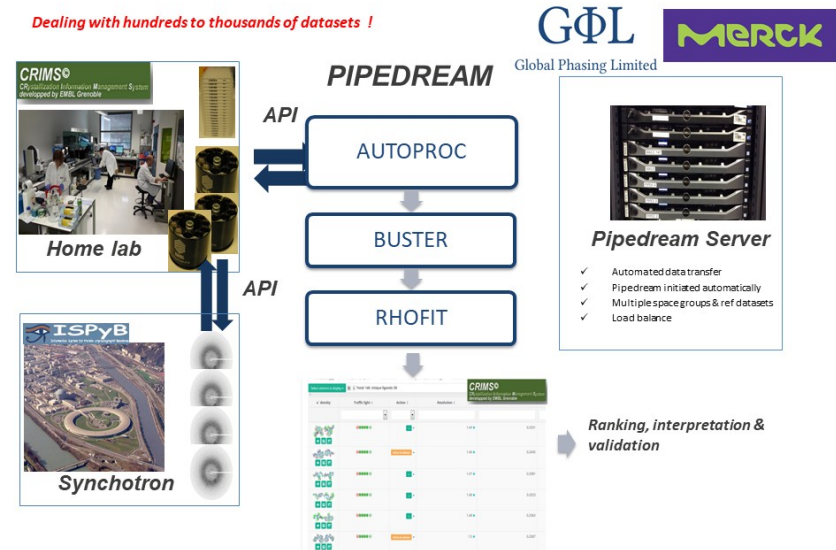
Created by: Marquez & Cipriani



Zander et al., Acta Cryst. (2016). D72, 454-466

Automated Data Flow: From protein to .pdb .mtz

Dealing with hundreds to thousands of datasets !



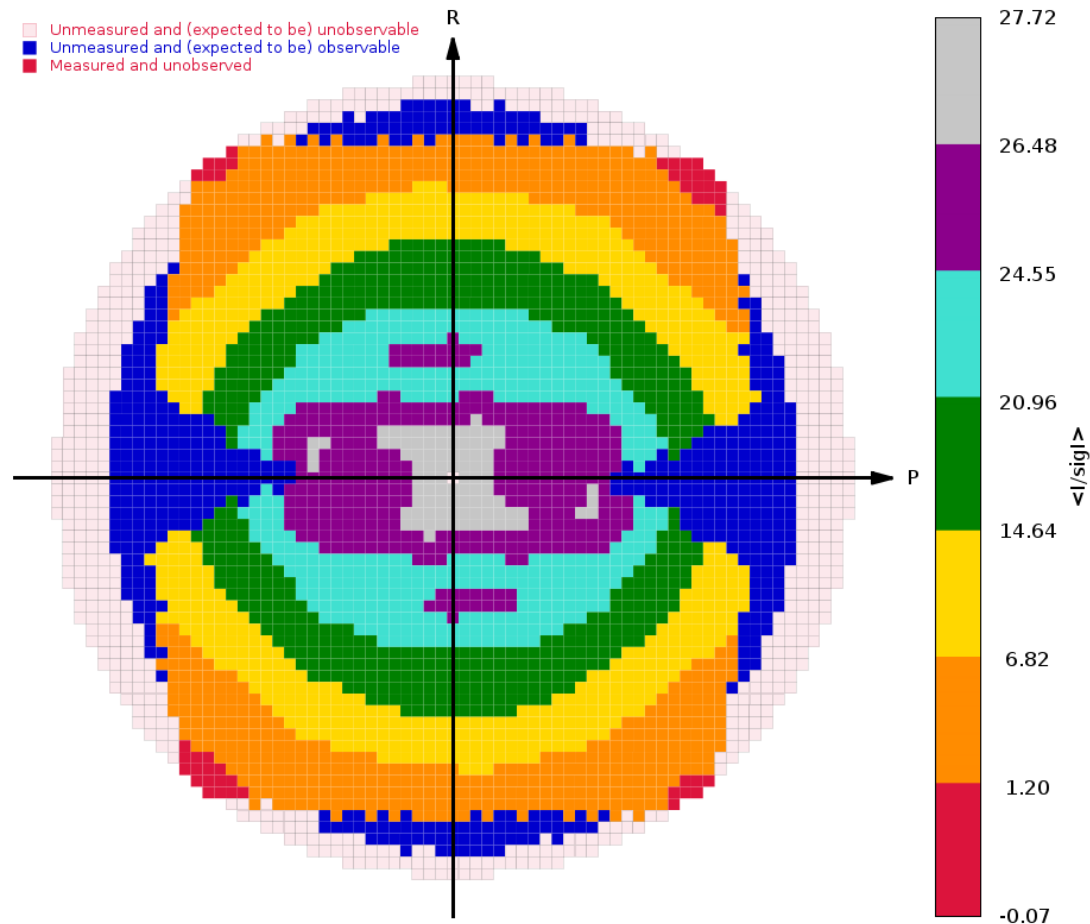
The Merck KGaA Collaboration: Validating Club Class Data Collection

- An HTX pilot project on FAK proposed by Merck KGaA provided an opportunity to validate Club Class data collection
- Specifically: to assess its impact on the effectiveness of a HT fragment-screening project compared to using datasets collected by standard protocols
- Two types of impact to be monitored:
 - effect of refining the reference (“apo” or DMSO) model against a specially collected Club Class dataset on the efficiency of ligand detection from ordinary datasets;
 - additional effect of using Club Class datasets for the ligand-soaked crystals as well.

The case of “1057”

- The first batch of FAK crystals soaked in a pre-selected library of Merck KGaA compounds yielded 135 datasets collected on the fully automated ESRF beamline MASSIF-1 by an “Economy Class” protocol typically consisting of a 167-degree sweep (for P1 crystals) on the single-axis RoboDiff goniometer.
- Code name “1057” will be used to designate one of these datasets and the compound into which the corresponding crystal had been soaked, and by extension later datasets collected on crystals soaked in the same compound.
- Pipedream brought up some difference density in the original 1057 dataset, but that density could not be fitted by RhoFit in terms of the known chemical structure of the compound used for soaking.

“Economy Class” dataset for 1057 (collected on MASSIF-1 at ESRF)



A PanDDA analysis of all 135 minimalistic MASSIF-1 datasets simultaneously (including 1057) *did not detect any binding for this ligand*

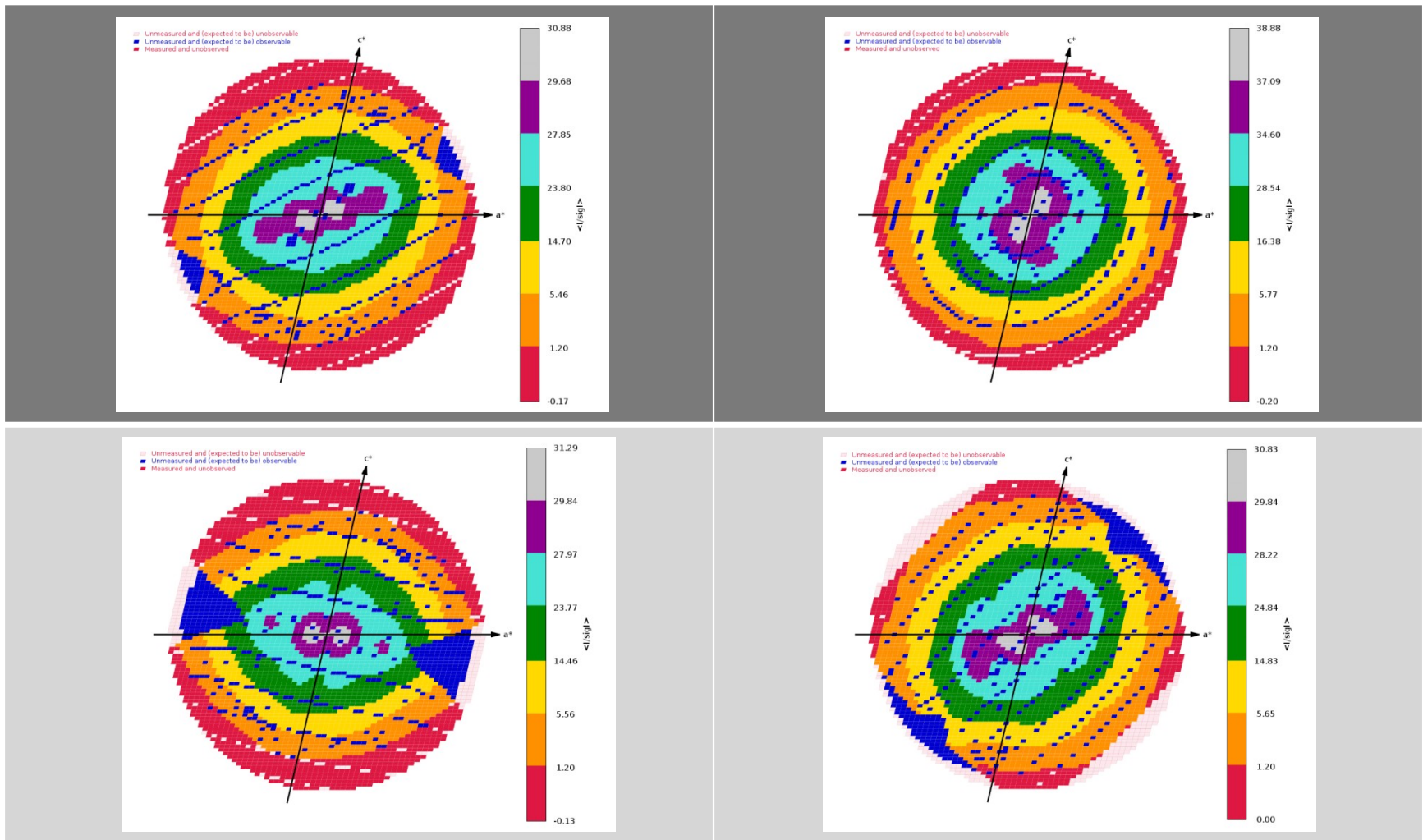
Dataset	Dataset Checks	RFree/RWork	Resolution	Reference RMSD	Map Uncertainty	Overall	Interesting Areas	Result
1051	✓ OK	0.261/0.24	✓ 1.438	✓ 0.099	✓ 0.315	✓ Analysed	✓ 1 Clusters	Z-Blobs Found
1054	✓ OK	0.257/0.228	✓ 1.647	✓ 0.166	✓ 0.345	✓ Analysed	✗	
1055	✓ OK	0.243/0.217	✓ 1.573	✓ 0.103	✓ 0.272	✓ Analysed	✗	
1056	✓ OK	0.263/0.246	✓ 1.428	✓ 0.113	✓ 0.359	✓ Analysed	✗	
1057	✓ OK	0.26/0.236	✓ 1.442	✓ 0.086	✓ 0.33	✓ Analysed	✗	
1058	✓ OK	0.257/0.237	✓ 1.46	✓ 0.145	✓ 0.35	✓ Analysed	✗	
1059	✓ OK	0.246/0.226	✓ 1.408	✓ 0.11	✓ 0.235	✓ Analysed	✓ 1 Clusters	Z-Blobs Found

***"If your experiment needs statistics,
you ought to have done a better
experiment"***

Ernest Rutherford

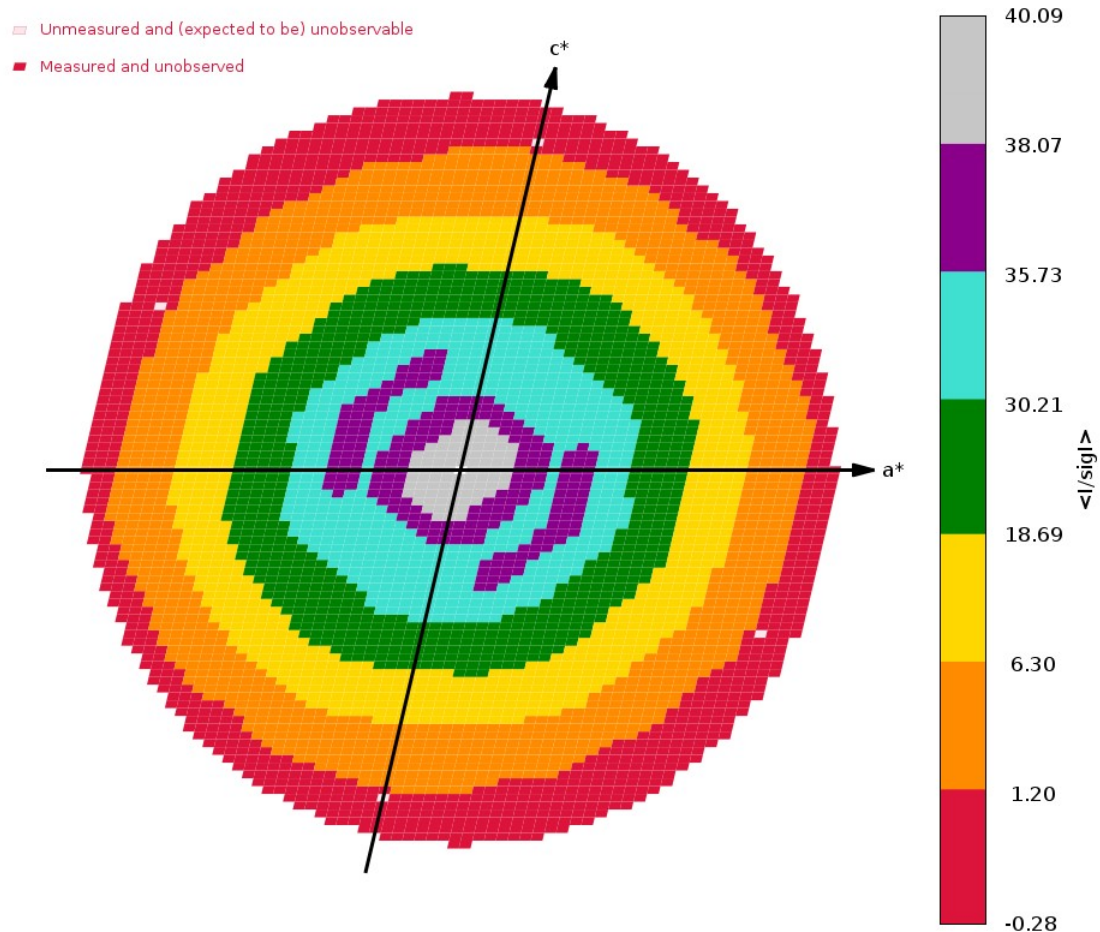
Alternative: go for high-quality individual datasets

The h0I STARANISO plots for 4 separate 360-degree sweeps on a P1 target



A four-sweep Club Class dataset (equal total dose to the initial MASSIF-1 dataset)

The h0I STARANISO plot for the merged 4-sweep dataset (4x360 degrees)



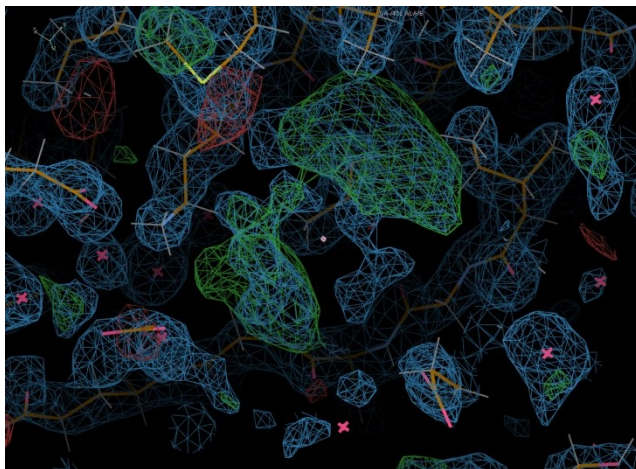
Comparative ligand screening results for 1057

GΦL

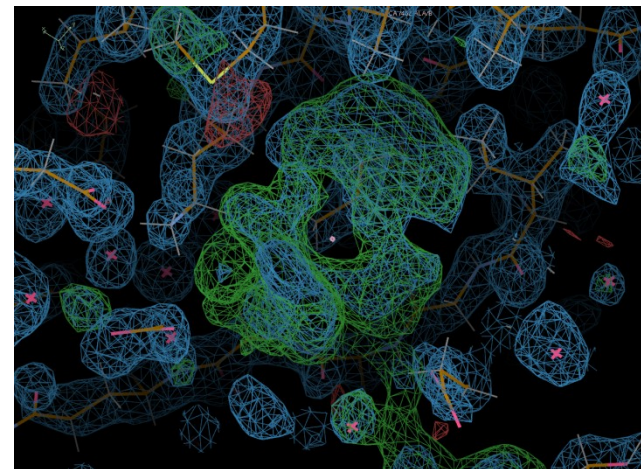
Global Phasing Limited

“Old” Model
(refined against
best MASSIF-1
dataset)

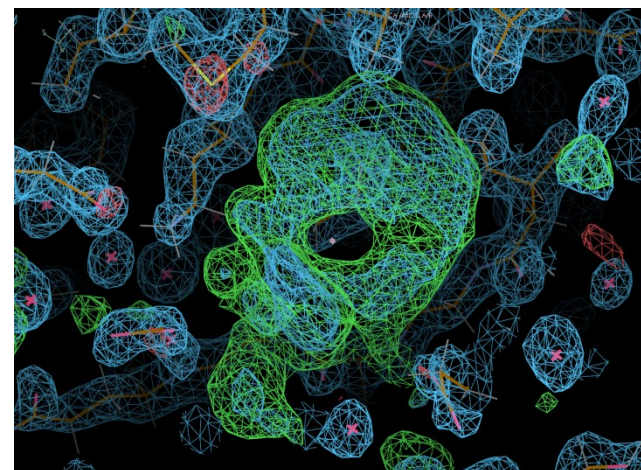
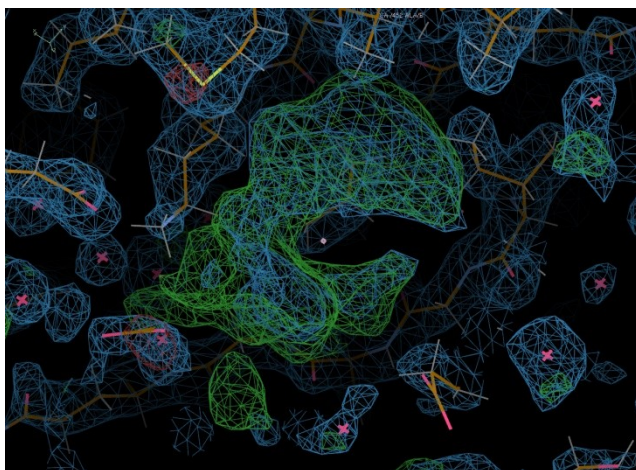
“Old” data for complex
(Standard MASSIF-1 protocol)



“New” data for complex
(Club-Class protocol on ID30-B)



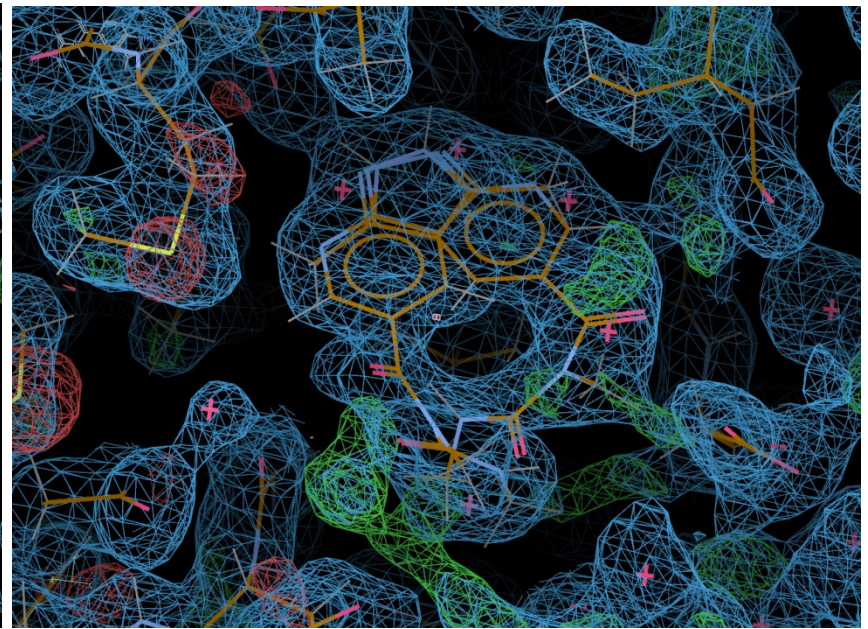
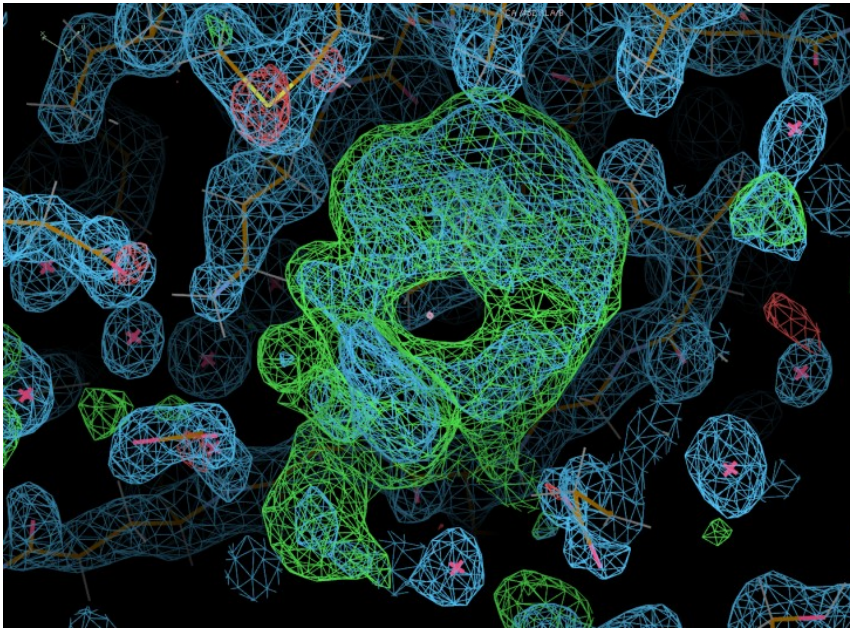
“New” Model
(refined against a
Club-Class
DMSO dataset)



Interpretation: two binding poses

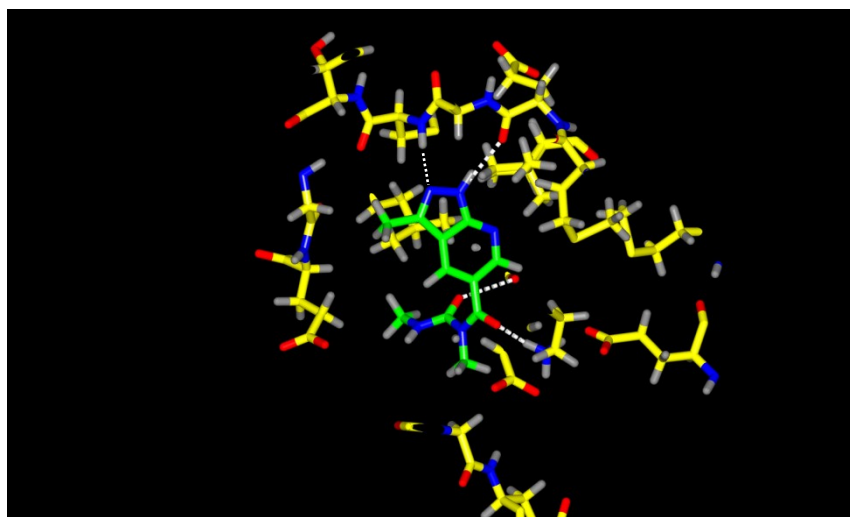
Difference density from new model
and new ligand-complex data

Interpretation in terms of two ligand
poses and residual waters from “apo”

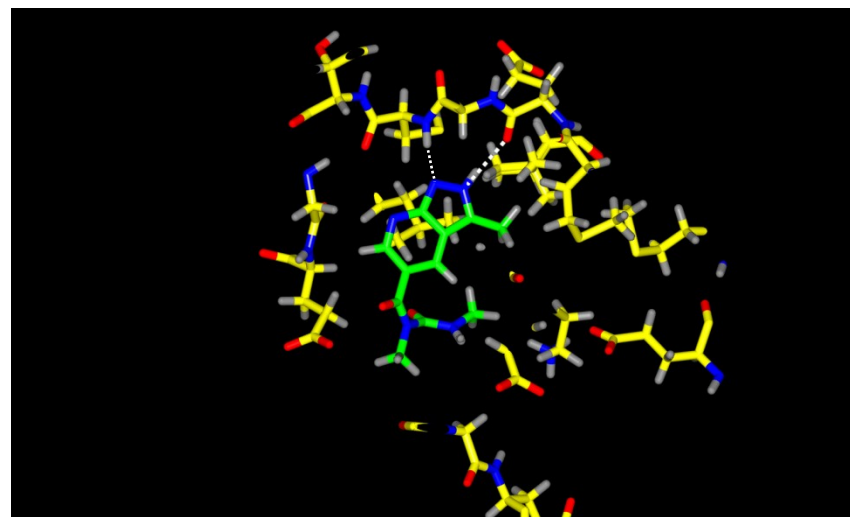


Final result for 1057

Binding mode A, tautomer 1



Binding mode B, tautomer 2



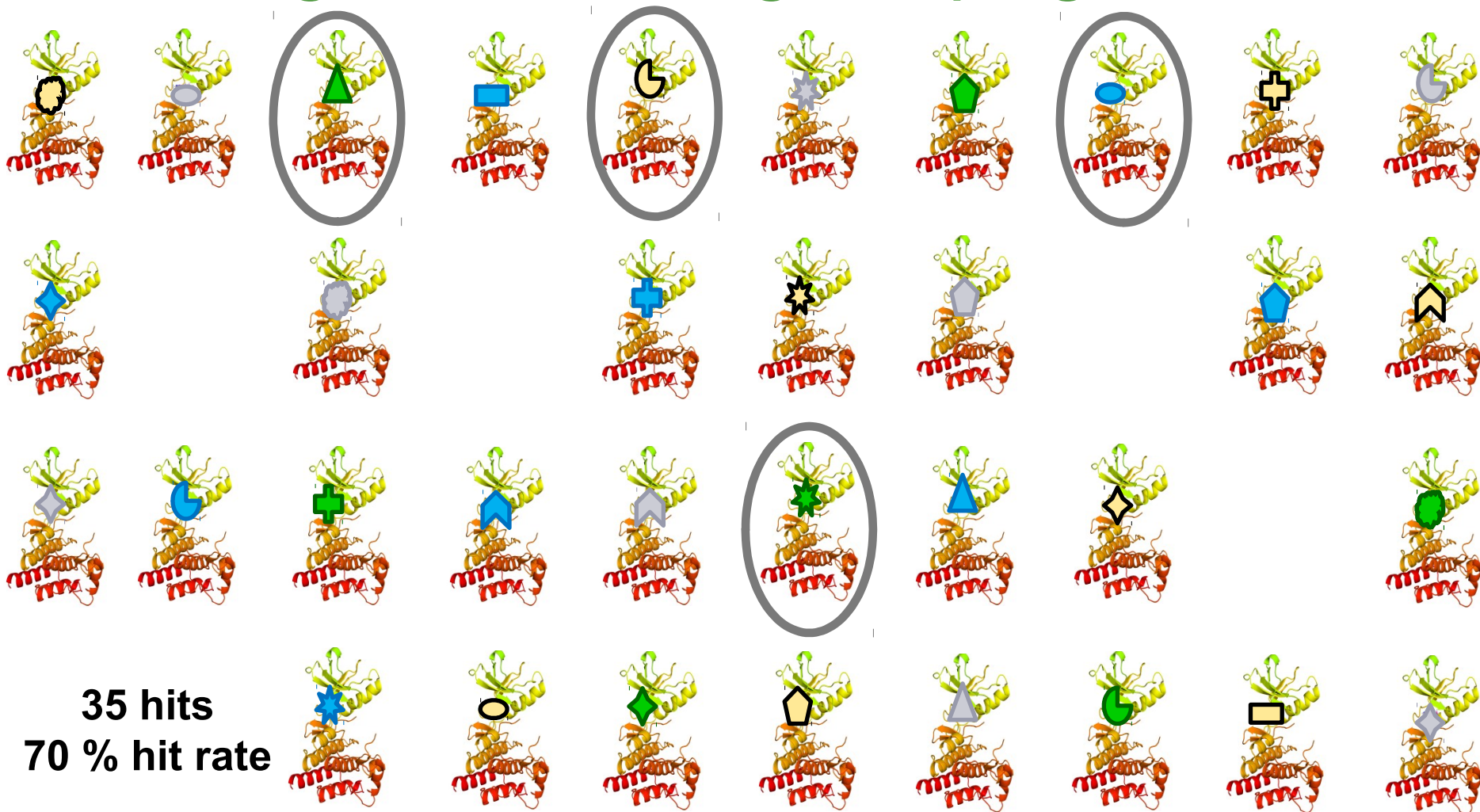
Conclusion:

- The use of Club Class datasets for both the (refinement of the) reference model and the 1057 putative ligand complex has not only established that something *is* bound, but has given a picture of the same fragment binding in two different poses involving two different tautomers for the pyrazole moiety.
- Another three compounds were rescued in the same way [data not shown]

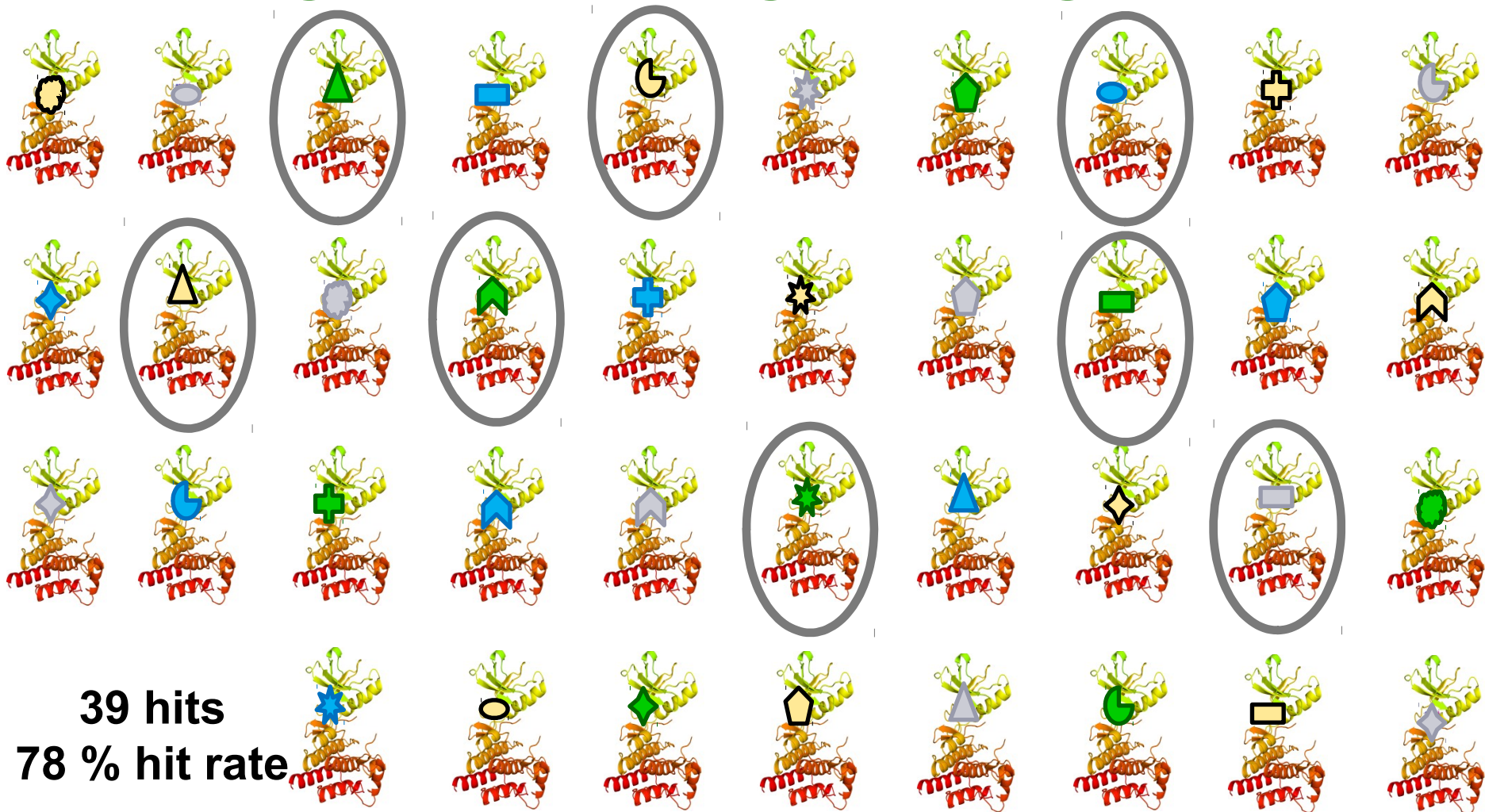
FAK Ligand screening campaign – results



FAK Ligand screening campaign – results



FAK Ligand screening campaign – results



39 hits

78 % hit rate

Returning to real-life complexities

- Importance of increasing the reliability of automated recentring between different orientations
 - **Pursuing better characterisation of the mini-kappa with ARINAX**
- Importance of connecting to all versions of MXCuBE and help re-unify it
 - **We are actively involved in the MXCuBE collaboration**
- Importance of being able to associate Club Class data collection requests to certain samples through a “diffraction plan”, and to access all results of workflow-triggered autoPROC via local synchrotron infrastructure:
 - **We are actively involved in the ISPyB collaboration**
- Importance to provide a framework for visualising and analysing the results of large numbers of Pipedream runs
 - **investigating the feasibility of using part of CRIMS for that purpose**