

# Global Phasing's ISPyB-related activities since the ELETTRA meeting

Rasmus Fogh, Peter Keller,  
Clemens Vonrhein, Gerard Bricogne

Global Phasing Ltd., Cambridge, UK

# Overview

- Continued deployment of autoPROC+STARANISO at synchrotrons
- Production of XML output to save anisotropic statistics in ISPyB
- In-lining of all plots and tables in summary.html to allow viewing through Web-based interfaces to ISPyB
- Dealing with rapid changes in the format of Eiger images produced by the DLS I04 beamline
- Throughout: participation in all ISPyB-DEV VCs
- Organising the workshop on Data Quality Metrics that was proposed at the end of the auto-processing session at the ELETTRA meeting.

# STARANISO

GΦL

Global Phasing Limited

## The STARANISO Server

*Anisotropy of the Diffraction Limit*

and

*Bayesian Estimation of Structure Amplitudes*



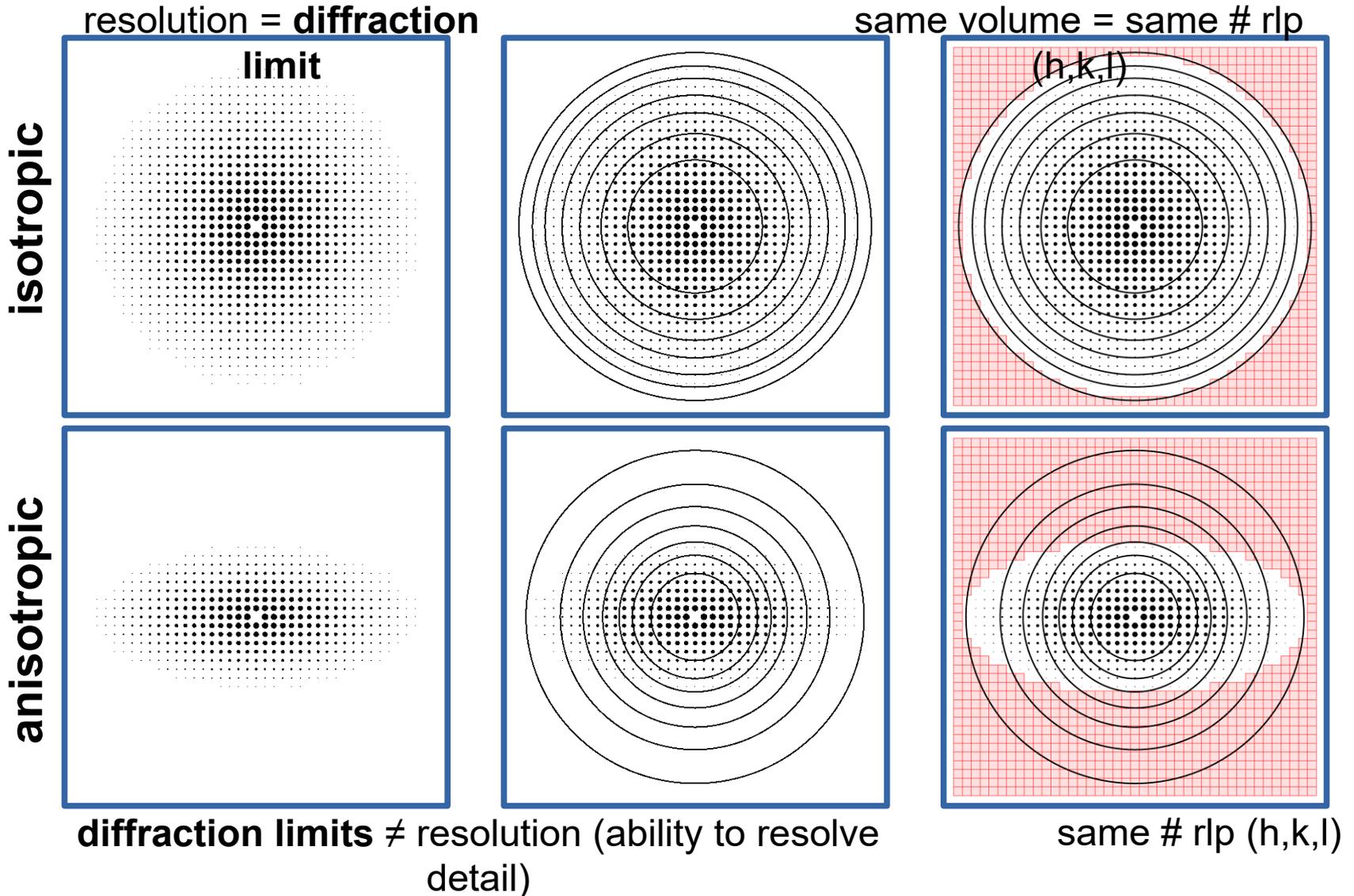
*Illicium verum*

main STARANISO server: [staraniso.globalphasing.org](http://staraniso.globalphasing.org)  
analyse deposited PDB datasets: [staraniso.globalphasing.org/cgi-bin/PDBpeep.cgi](http://staraniso.globalphasing.org/cgi-bin/PDBpeep.cgi)

Remember: anisotropy means “not isotropic” (an ellipsoid is only a useful approximation to simplify description of actual anisotropy).

**Tickle, I.J.**, Flensburg, C., Keller, P., Paciorek, W., Sharff, A., Vornrhein, C., Bricogne, G. (2018). STARANISO. Cambridge, United Kingdom: Global Phasing Ltd.

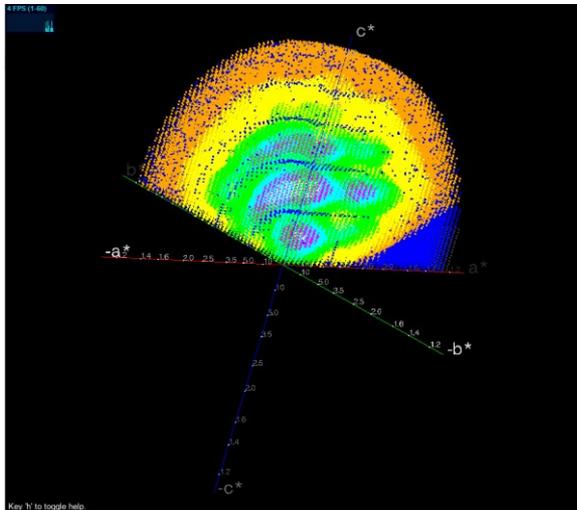
# Resolution, diffraction limit and equal-observation-number binning



# 3D WebGL visualisation in STARANISO

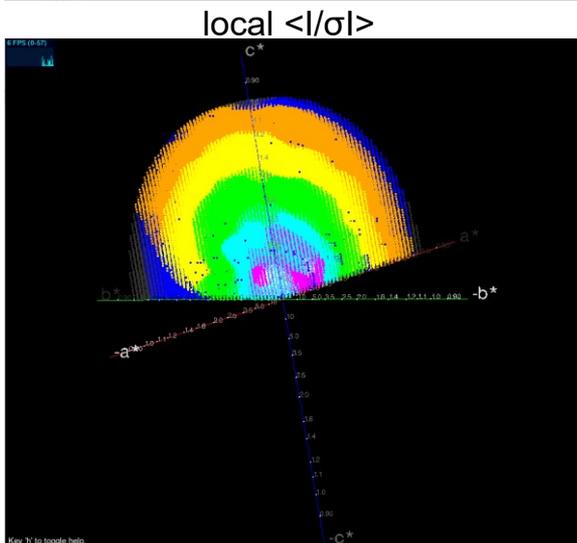
<http://staraniso.globalphasing.org/cgi-bin/PDBpeep.cgi>

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



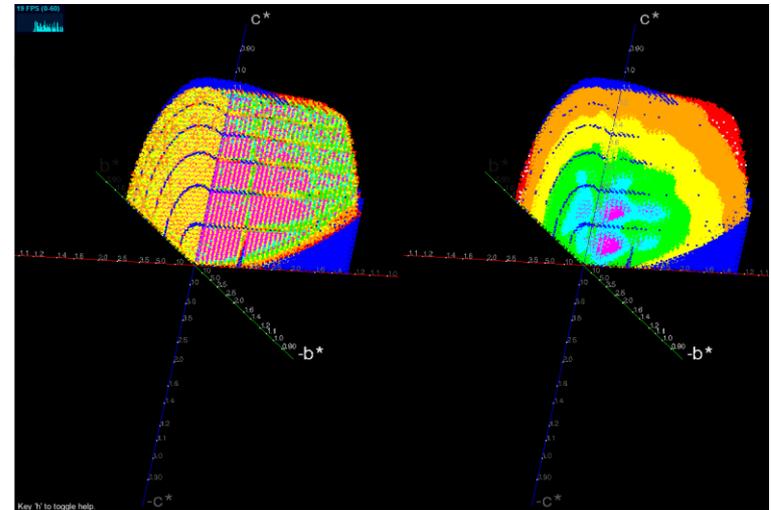
4YUO P<sub>2</sub><sub>1</sub><sub>2</sub><sub>1</sub><sub>2</sub><sub>1</sub>  
 $a^* \sim 7^\circ$   
single sweep

beam centre in  
module gap



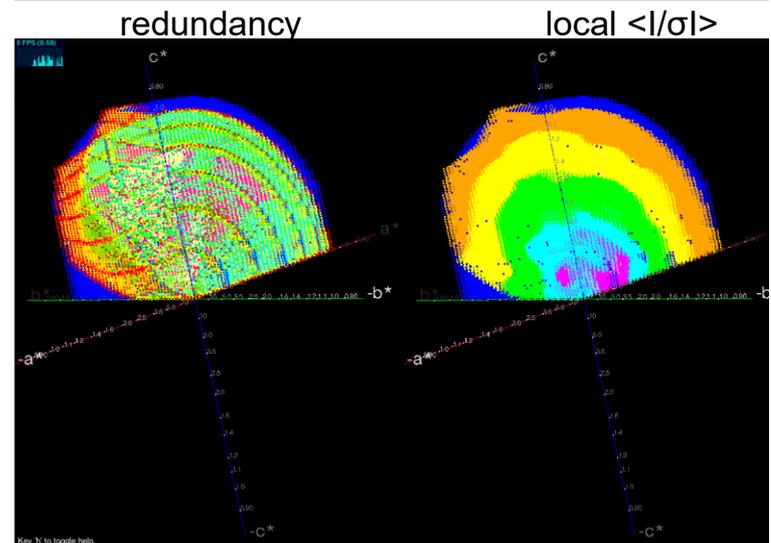
local  $\langle I/\sigma \rangle$

merged data



redundancy

unmerged data



local  $\langle I/\sigma \rangle$

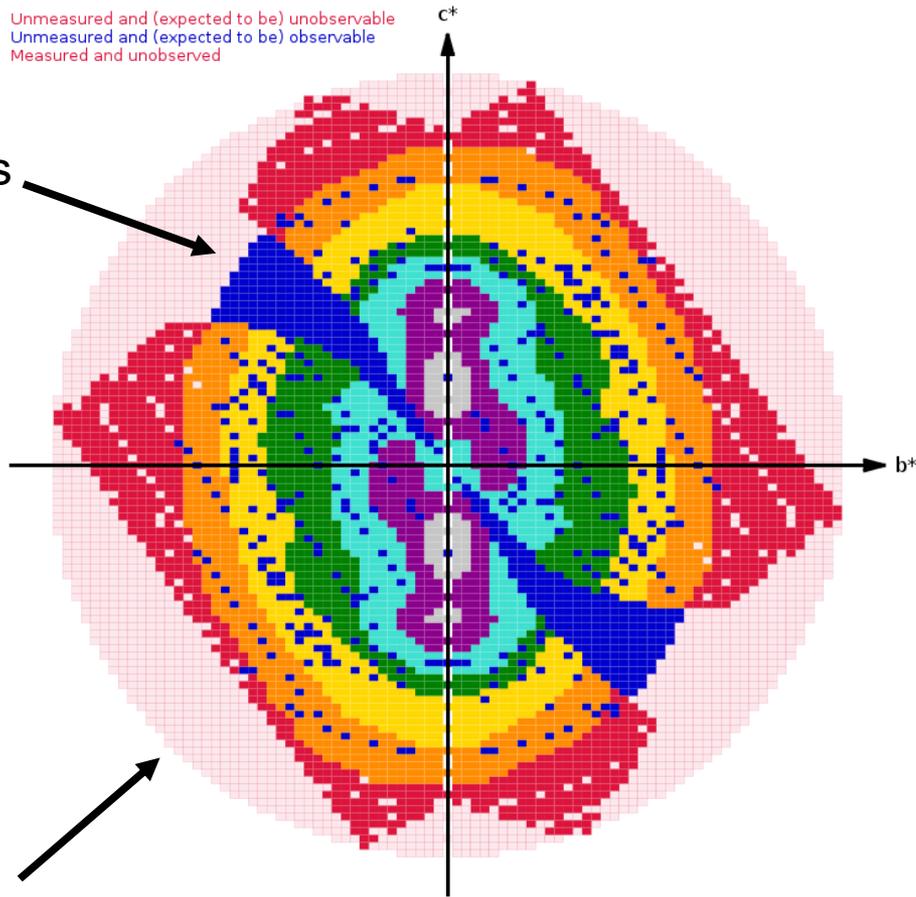
4ZC9 P<sub>2</sub><sub>1</sub><sub>2</sub><sub>1</sub><sub>2</sub><sub>1</sub>  
 $b^* \sim 14^\circ$   
2 sweeps  
(distance)

# autoPROC: STARANISO plots

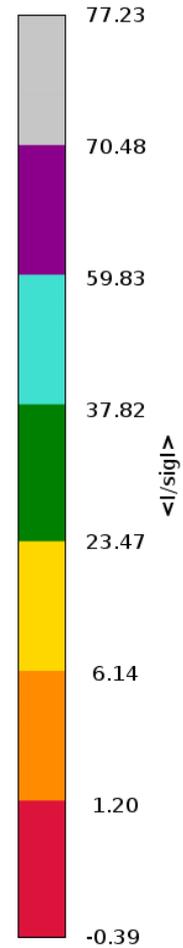
rec. unit cell

- Unmeasured and (expected to be) unobservable
- Unmeasured and (expected to be) observable
- Measured and unobserved

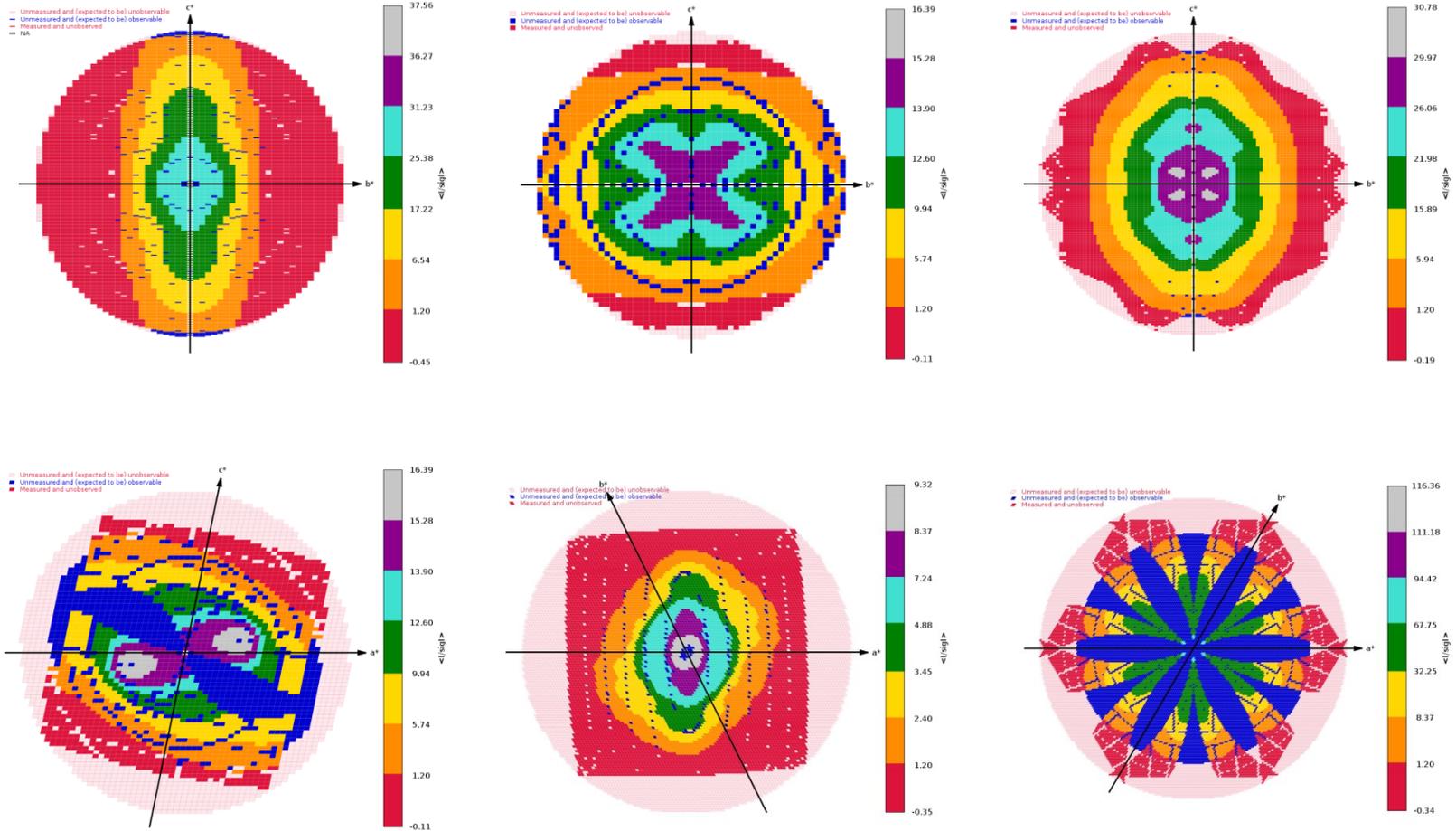
Diffraction limits observable



Diffraction limits recorded



# STARANISO within autoPROC – 2D plots



# Breaking the mould

- **Somehow autoPROC+STARANISO is “disruptive”!**
  - **Pro:** the coolest thing to be nowadays ;-)
  - **Con:** difficulties with the **Procrustean bed** of current archiving/deposition software and of pervasive thinking habits and practices caused by quantities such as resolution cut-off and completeness being thought of as single-valued, while they necessarily become multi-valued when anisotropy is present
- **The extra statistics produced by the STARANISO analysis** have to be not just archived but also presented and made available for decision-making.
- **Beyond statistics, the 3D picture of the distribution** of signal/noise produced by the STARANISO server is very information-rich and has been promoted by Bernhard Rupp as tomorrow’s (why not today’s?) replacement for Table I (see <https://lnkd.in/etSFFAZ>)

## Against Method: Table 1—Cui Bono?

Bernhard Rupp<sup>1,2,\*</sup>

<sup>1</sup>k.-k.Hofkristallamt, San Diego, CA 92084, USA

<sup>2</sup>Division of Genetic Epidemiology, Medical University Innsbruck, Schöpfstraße 41, Innsbruck, Tyrol 6020, Austria

\*Correspondence: [bernhard.rupp@i-med.ac.at](mailto:bernhard.rupp@i-med.ac.at)

<https://doi.org/10.1016/j.str.2018.04.013>

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.”

### Introduction

Cui bono?—To whose benefit?—is an important question for legal proceedings and an equally relevant one for the dissemination of experimental information in scientific publishing. In principle, a useful item in the experimental section of a scientific pub-

only an *upper limit* for the overall achievable model quality, with most of the data metrics only measures of precision but not of accuracy.

### Obsolescence

Table 1 is essentially a remnant from the pre-deposition era when

# Reciprocal Space Data Visualization

Present.....and past (1913)

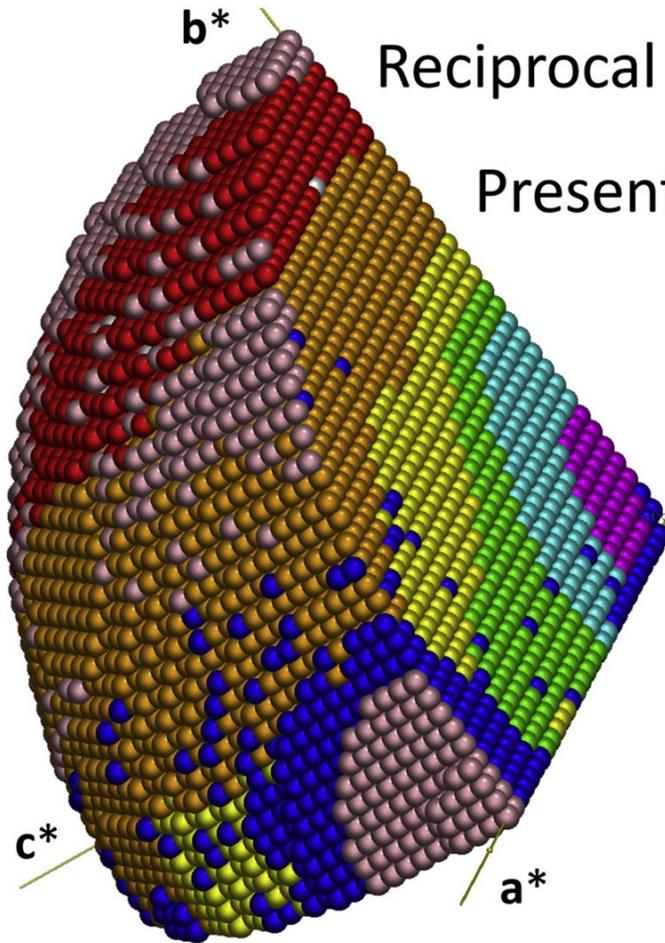


Table I.  
KCl.

| $h \rightarrow$ | 0 | 1 | 2 | 3 | 4 | 5 |
|-----------------|---|---|---|---|---|---|
| $k \downarrow$  |   |   |   | * | * |   |
| 0               |   |   |   | * | * |   |
| 1               |   |   |   | * | * |   |
| 2               |   |   | * | * | + |   |
| 3               | * | * | * | + |   |   |
| 4               | * | * | + |   |   |   |
| 5               |   |   |   |   |   |   |

# Hic et nunc: what do we do about Table 1?

```
Spacegroup name      P1
Unit cell parameters 45.335 50.262 66.953 74.299 88.909 89.972
Wavelength           0.97625 A
```

Diffraction limits & principal axes of ellipsoid fitted to diffraction cut-off surface:

```
2.425      0.7195  0.5966  0.3556   0.580  a* + 0.533  b* + 0.616  c*
1.985     -0.6627  0.7429  0.0947  -0.584  a* + 0.726  b* + 0.364  c*
2.848     -0.2076 -0.3038  0.9298  -0.165  a* - 0.268  b* + 0.949  c*
```

Number of active ice-rings within this resolution range = 1

Criteria used in determination of diffraction limits:

-----  
local(I/sigI) >= 1.20

|                                      | Overall | InnerShell | OuterShell |
|--------------------------------------|---------|------------|------------|
| -----                                |         |            |            |
| Low resolution limit                 | 48.386  | 48.386     | 2.305      |
| High resolution limit                | 2.038   | 6.747      | 2.038      |
| <br>                                 |         |            |            |
| Rmerge (all I+ & I-)                 | 0.099   | 0.043      | 0.614      |
| Rmerge (within I+/I-)                | 0.087   | 0.038      | 0.521      |
| Rmeas (all I+ & I-)                  | 0.117   | 0.051      | 0.726      |
| Rmeas (within I+/I-)                 | 0.123   | 0.054      | 0.736      |
| Rpim (all I+ & I-)                   | 0.062   | 0.028      | 0.384      |
| Rpim (within I+/I-)                  | 0.087   | 0.038      | 0.521      |
| Total number of observations         | 67580   | 3256       | 3429       |
| Total number unique                  | 19451   | 973        | 973        |
| Mean(I)/sd(I)                        | 5.7     | 14.0       | 1.4        |
| Completeness (spherical)             | 53.5    | 97.2       | 8.7        |
| Completeness (ellipsoidal)           | 86.7    | 97.2       | 59.4       |
| Multiplicity                         | 3.5     | 3.3        | 3.5        |
| CC(1/2)                              | 0.996   | 0.997      | 0.768      |
| <br>                                 |         |            |            |
| Anomalous completeness (spherical)   | 52.7    | 95.7       | 8.6        |
| Anomalous completeness (ellipsoidal) | 85.2    | 95.7       | 58.9       |
| Anomalous multiplicity               | 1.8     | 1.7        | 1.8        |
| CC(ano)                              | -0.110  | -0.093     | -0.219     |
| DANO /sd(DANO)                       | 0.520   | 0.471      | 0.553      |

# Table 1? What about ISPyB first?

- The new statistics produced by STARANISO within autoPROC have the potential of **better characterising dataset quality**, but effort is needed to get them **accepted and accommodated by synchrotron data management software**
- **Next: same thing with the PDB!**
- **The experience of our “power users”** in evaluating the usefulness of these statistics and of metrics based on them will be critical in enabling us to overcome the obstacles involved: there must be more direct communication with them.
- **Our outlook:** these metrics will be used to plan and supervise better data collection experiments.

# Workshop on Data Quality Metrics

## ESRF, April 9-11, lunchtime to lunchtime

The purpose of this workshop is to carry out a comprehensive review of the current practices for selecting and displaying a variety of data quality metrics pertaining to the results output by pipelines deployed at various synchrotrons for the auto-processing of diffraction images.

The intended tone of the workshop is to be constructively iconoclastic. It will look beyond the present state of affairs and seek to identify a small number of coordinated action points between different groups of protagonists to reduce latency in the process of translating new scientific insights into improved user-facing practical procedures for quality assessment, visualisation, validation and archiving.

For this purpose, the workshop will bring together (1) active processing software and data analysis specialists; (2) programmers involved in developing and operating auto-processing pipelines as well as databases of experiments and results (e.g. ISPyB) and interfaces for presenting summaries of those results to users (e.g. EXI, SynchWeb); (3) experienced "power users" (e.g. from the pharma industry) who are among the main consumers of this information; and finally (4) deposition and archiving developers from the PDB(e).

# In the meantime have fun with ...



## **WebGL**. interactive 3-D reciprocal lattice and anisotropy viewer

- [About ANISOTROPY](#)
- [STARANISO main server](#)
- **About this server**

**PDBpeep** is intended only for a rough quick visualization ('peep') of the reciprocal lattice colored by the local mean  $I/\sigma(I)$  and the [Debye-Waller factor](#) for intensity or amplitude diffraction data deposited in the [Protein Data Bank](#). It is NOT our intention to provide any outputs, other than textual and graphical descriptions of the anisotropy, otherwise we would be ignoring our own advice to run STARANISO only on data to which no cut-offs have previously been applied! Users who really want to perform such a 'correction' can always download the files from the [PDB servers](#) themselves and submit them to the [STARANISO main server](#) (but be it on your own head should you choose to do so!).

- **Submit PDB ID:**

**NOTE:** the 4-character PDB ID entered in the box above MUST have intensities (unmerged or merged) and/or observed structure amplitudes (both with standard uncertainties) deposited.  
Calculated structure factors, map coefficients, phases, free R flags *etc.* are all ignored if present.

- **Submission counter**

|      | Jan | Feb | Mar | Apr | May | Jun | Jul | Aug | Sep | Oct | Nov | Dec | Σ   |
|------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| 2018 | 0   | 0   | 368 | 154 | 65  | 179 |     |     |     |     |     |     | 766 |

This software has been developed at [Global Phasing Ltd.](#) thanks to funding from the Global Phasing Consortium.

The following have all contributed to the development of the STARANISO servers in various ways too numerous to attribute individually, whether it be ideas, code, tools or testing:

|                 |                  |
|-----------------|------------------|
| Ian J. Tickle*  | Andrew Sharff    |
| Claus Flensburg | Oliver Smart     |
| Peter Keller    | Clemens Vonrhein |
| Wlodek Paciorek | Gérard Bricogne  |

\*contact: [staraniso@globalphasing.org](mailto:staraniso@globalphasing.org)



# Acknowledgments

- Ian Tickle (Global Phasing)
- Olof Svensson, Stephanie Monaco (ESRF)
- The ISPyB developers for interesting VCs