Data management and processing pipelines at DLS

Alun Ashton
Group Leader Data Analysis Software
Diamond Light Source

- SFX XFEL UK Hub
- EM Facility
Life Sciences @ DLS

Cell/Tissue
- B22 Infrared
- B24 Cryo X-ray microscopy
- I08 X-ray scanning microscope
- I13 X-ray tomography & coherent diffraction
- I18, I20, B18, I14 X-ray spectroscopy

Solution
- I22/B21 SAXS
- B23 CD
- Spectroscopy village

Crystalline
- MX village
  (I02, I03, I04)
  (I24, I04-1)
  (I23, VMX)
- XFEL Hub

Electron Microscopy Facility
- National facility for EM
- Life and Physical sciences
Diamond’s Role in Structural Cell Biology

• Provide and develop techniques to cover all resolution ranges
  • Diffraction, X-ray and electron microscopy, x-ray spectroscopy, imaging....
• Focus for an integrated approach to structural biology at RAL
• to initiate new approaches/techniques
An integrated structural biology platform for the UK

- Cell Biology
- OPPF-UK
- MPL
- RC@H

- UK XFEL
- HuB@Diamond

- Diamond Beamlines: Macromolecular Crystallography, Scattering, X-ray spectroscopy

- Computational environment / CCP4

- X-ray imaging
- Fluorescence microscopy (CLF/STFC & DLS)
- Cryo-EM/ET
Data analysis Pipeline for Macromolecular Crystallography

- Automated data collection and robotics
- Remote control for users
- Full LiMS system and interfaces
- Fully automated structure solution pipelines, in 80% of cases

Change Sample

Fast Data Processing: Fast_DP

Archive Data icat.diamond.ac.uk

Diff Map: DIMPLE

Experimental Phasing: Fast_EP

Molecular Replacement MR Bump

Rerunning Jobs

Data Processing xia2 – 2d and 3d

~ 10-30 minutes
SynchLink iOS App.
BioSAXS Experiment
Automated BioSAXS Reduction
ISPyB information Flow: Display of Results

Data tracking and status checking of data acquired:

- Automatic plot generation of Scattering curves, Kratky, Guinier and P(r) plots

<table>
<thead>
<tr>
<th>Macromolecule</th>
<th>Concentration</th>
<th>Scattering</th>
<th>Kratky</th>
<th>Guinier</th>
<th>P(r)</th>
<th>Frames (Average/Total)</th>
<th>Gaurier</th>
<th>Gaurier</th>
<th>Dmin</th>
<th>Dmax</th>
<th>Volume</th>
<th>PDI (Vol) Yol</th>
<th>Yol est.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Iso</td>
<td>5.86 mg/ml</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- Display of 1D curves, averaged curves and subtracted curves
- 3D Models displayed using WebGL
- Concentration Effects

A prototype version of ISPyB for BioSAXS (ISPyBB) database is now available at BM29@ESRF and is being ported to P12@EMBL
Work on other techniques ongoing.
Diamond has a policy of, where feasible, to standardise on file formats, the choice being NeXus/HDF5.

**Green:** predominantly using NeXus.

**Orange:** Mixed NeXus and other formats or considering NeXus in the next 12 months.

Files can be generated by Detector, EPICS or Data Acquisition.
Archive and Reprocessing Service

- We have a copy of > 99% of all Diamond data and data processed at Diamond on tape, 80% of which available from a web site https://icat.diamond.ac.uk and as a mount point on DLS disk and in Eclipse. Ability to publish data with a DOI by 2013.
- Remote desktop service for reprocessing data at Diamond.
  - Using off the shelf NX solution for remote desktop www.nomachine.com
  - 4 x 8 CPU clustered server.
Current status/volumes in Diamond
ICAT = 285,198,074 files

This image shows the diffraction pattern of an RNase 4 crystal at 1.6Å, from Prof K. Ravi Acharya's group (at University of Bath) - the image that took DLS total data catalogued and archived > 1 Petabyte.
Beamline Computing/Software Support

- Data Analysis Group: – Alun Ashton
- Data Acquisition Group: – Paul Gibbons
- Scientific Computing Team: - Greg Matthews
- Beamline Controls Group: - Nick Rees
- User Office development team: – Bill Pulford, Ben Peacock