

EDNA

A tool for online data analysis

and

A collaborative effort

**Olof Svensson
EDNA Project Manager
ESRF Data Analysis Unit**

Overview

- Online data analysis
- The EDNA MXv1 Characterisation Application
 - Data collection taking into account radiation damage
 - The first EDNA release : June 4th 2009
- The EDNA Project / Framework
 - Project management
 - Data model
 - Modularity / plugins
 - The testing framework
- Future developments
 - MXv2 / Tomography
- Questions? Don't hesitate to interrupt me!

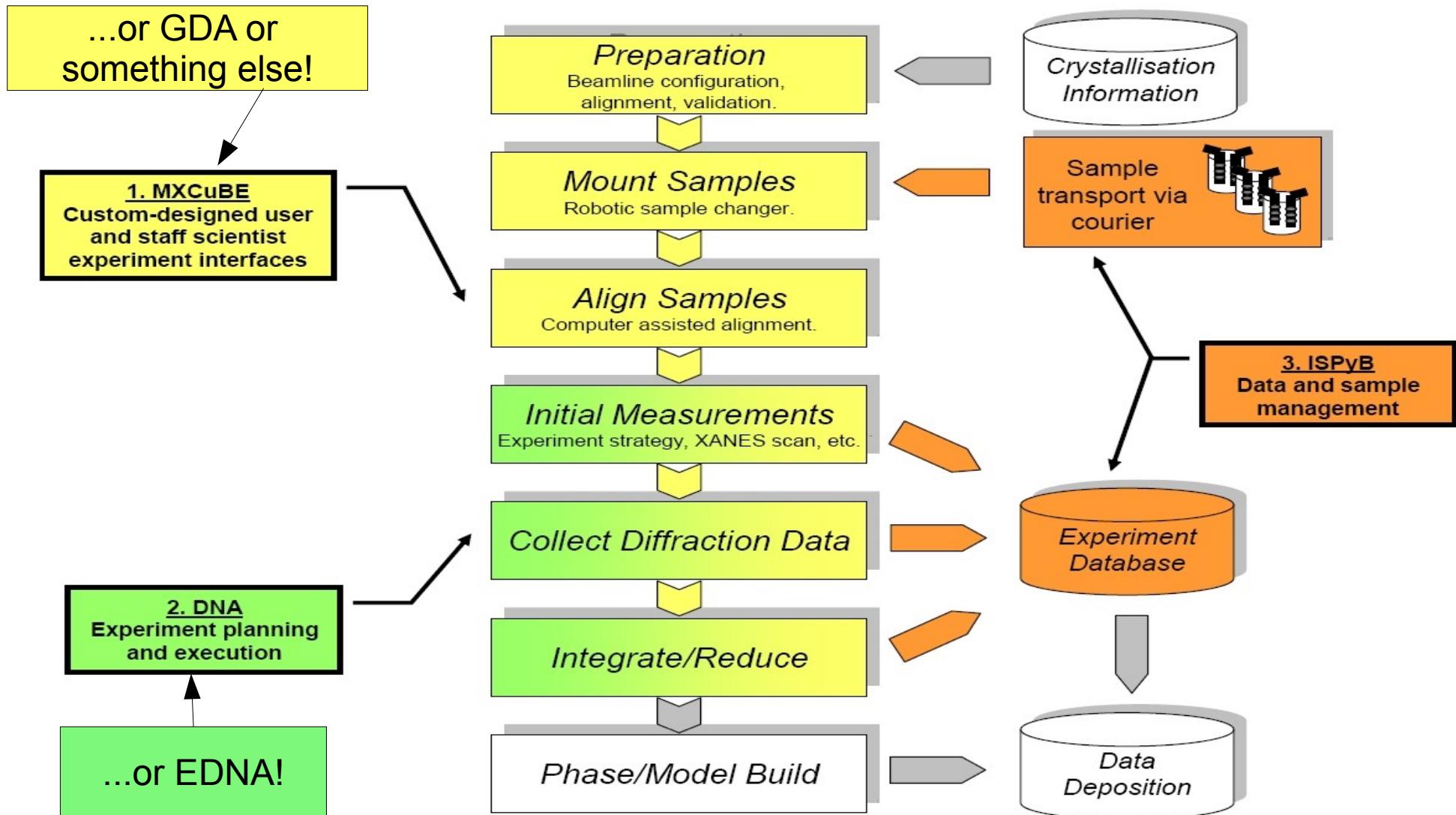
Online Data Analysis (1)

- Automation
 - Example MX : Data Collection Pipeline (DCP)
- Grid / Batch processing
 - Fast data processing ⇒ parallel / distributed execution

Online Data Analysis (2)

- Data Management
 - Lims
- Modular
 - It must be easy to change the work flow / scientific programs
 - Data Model
- Robust - code reliability
 - If ODA fails once, the user won't try again...
 - Efficient error tracking system
- Collaboration
 - Avoid re-inventions of the wheel
 - Project management
 - Documentation, Use Cases

MX Automation - the Data Collection Pipeline: Full automation from sample loading to reduced (integrated and scaled) data

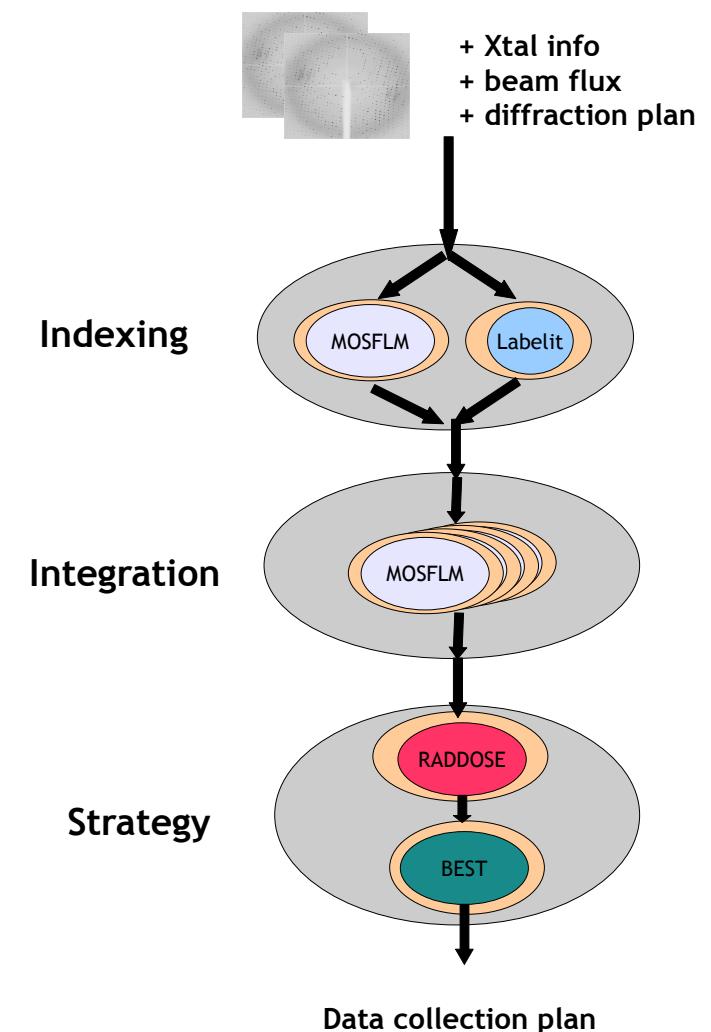


What is EDNA?

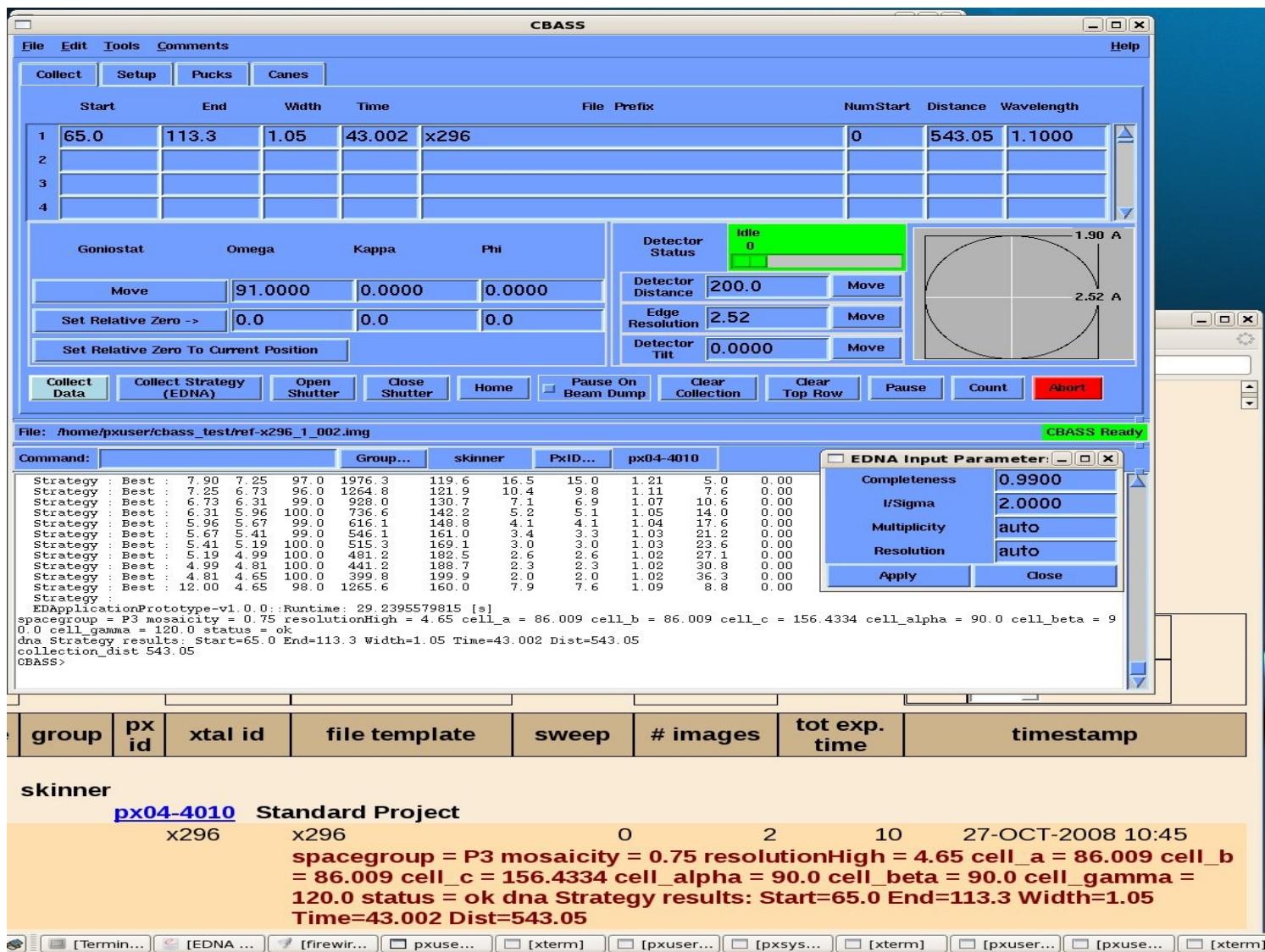
- EDNA is an international collaborative project between several institutes and synchrotron facilities.
- Developed on the foundation of the project automateD collectioN of datA (« DNA », www.dna.ac.uk)
- Designed to be a framework for Online Data Analysis of X-ray experiments

EDNA MXv1 Characterisation

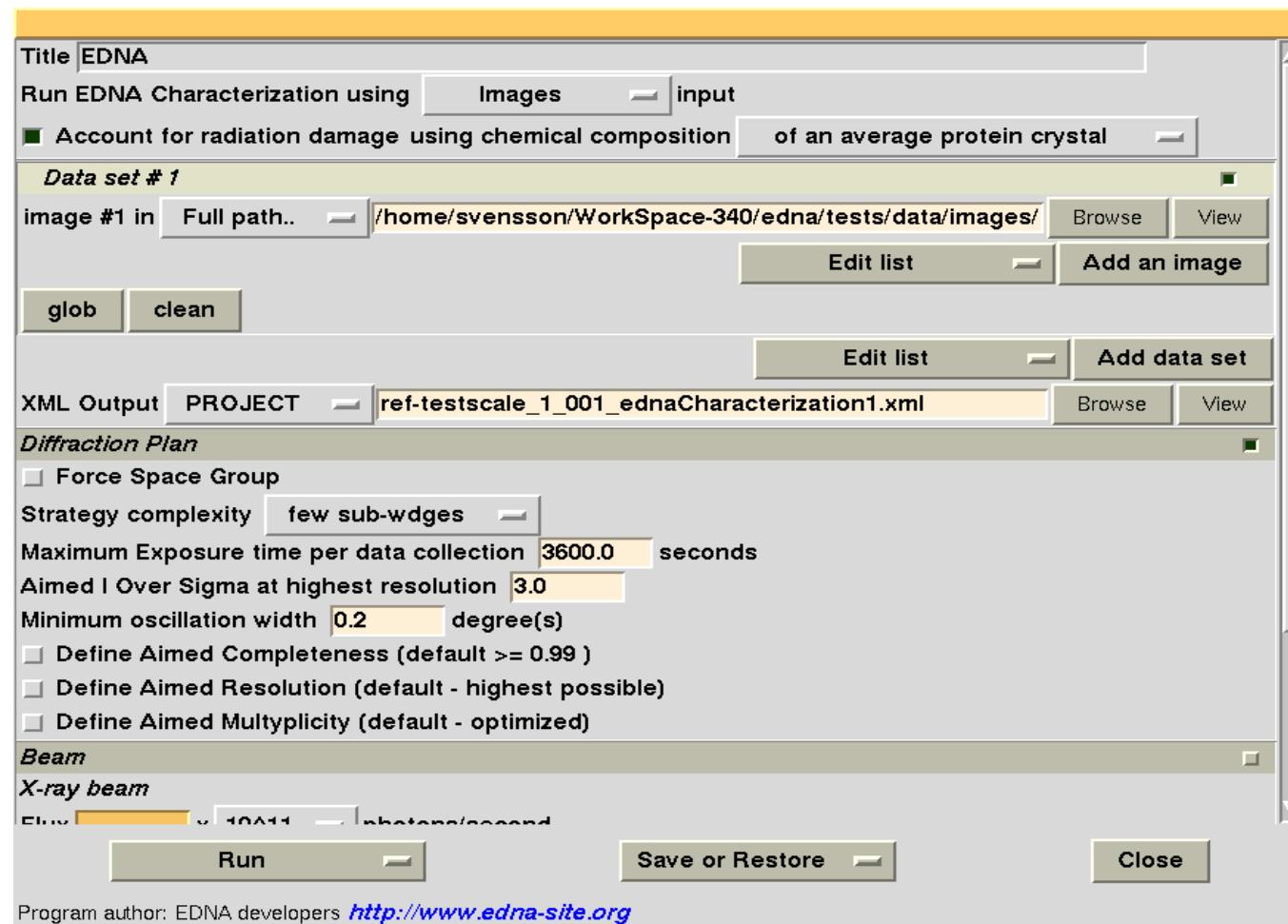
- MX sample characterisation taking into account radiation damage
 - Indexing using MOSFLM or Labelit
 - Parallel integration of reference images
 - If flux +beamsize + chemical composition: RADDOSE for estimating radiation damage
 - BEST strategy calculation
 - taking into account radiation damage
 - multi-subwedge data collection strategies



NSLS CBASS GUI



EDNA MXv1 Characterisation ccp4i GUI



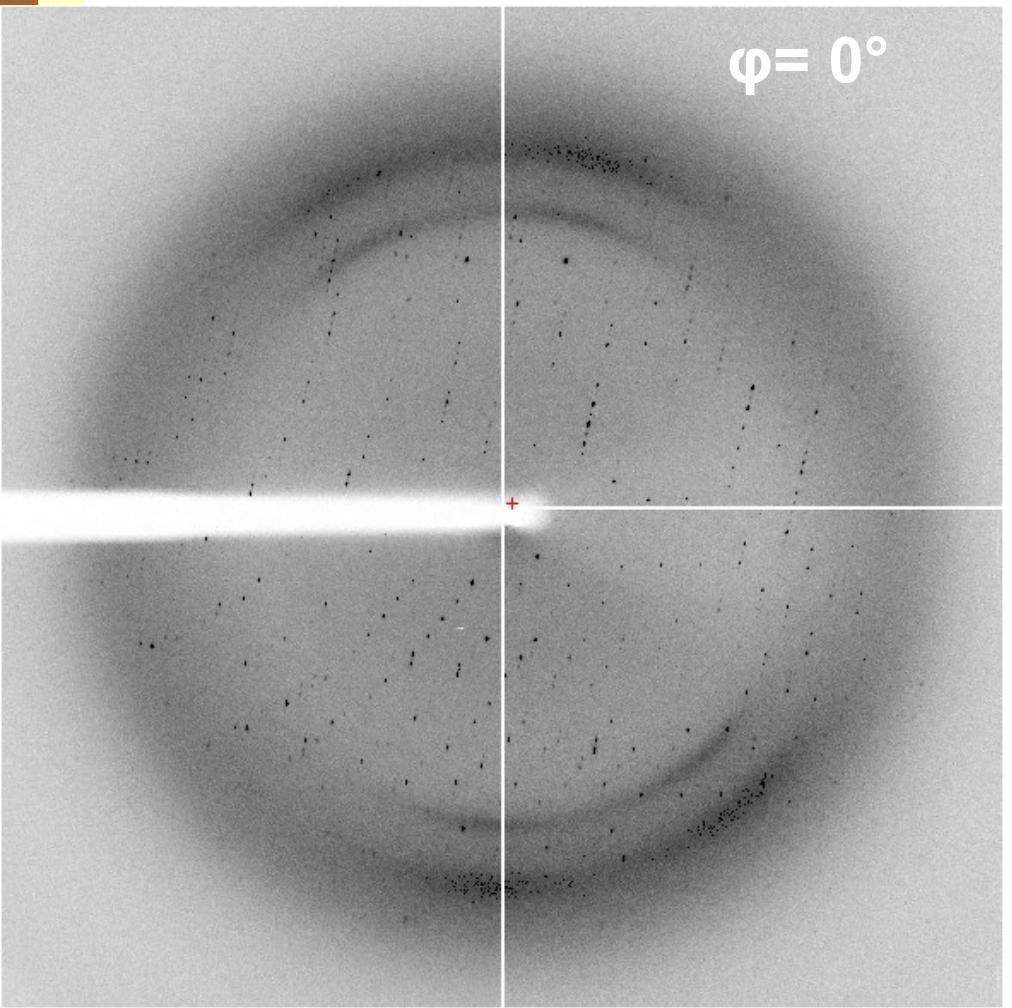
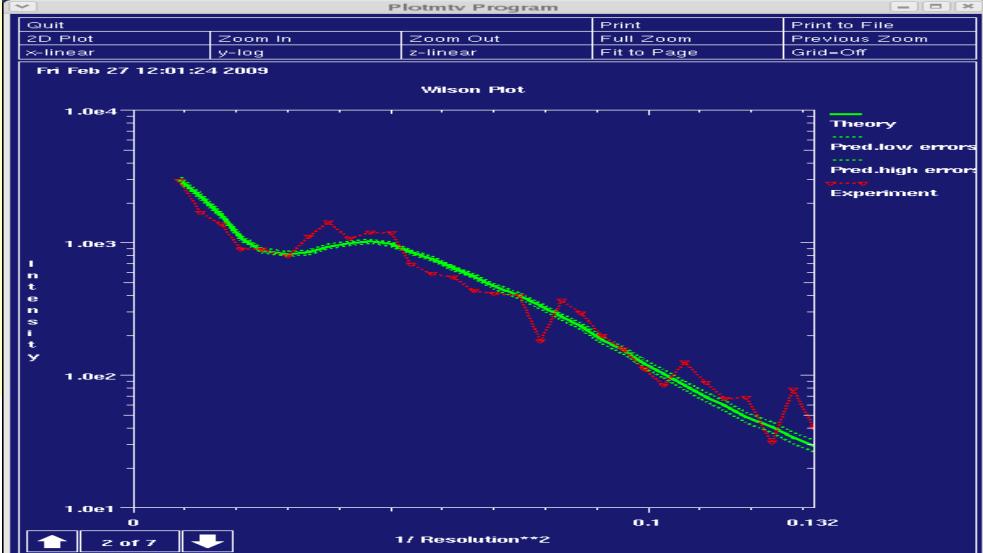
ID14-2, ESRF

Initial images for Survival E protein crystal

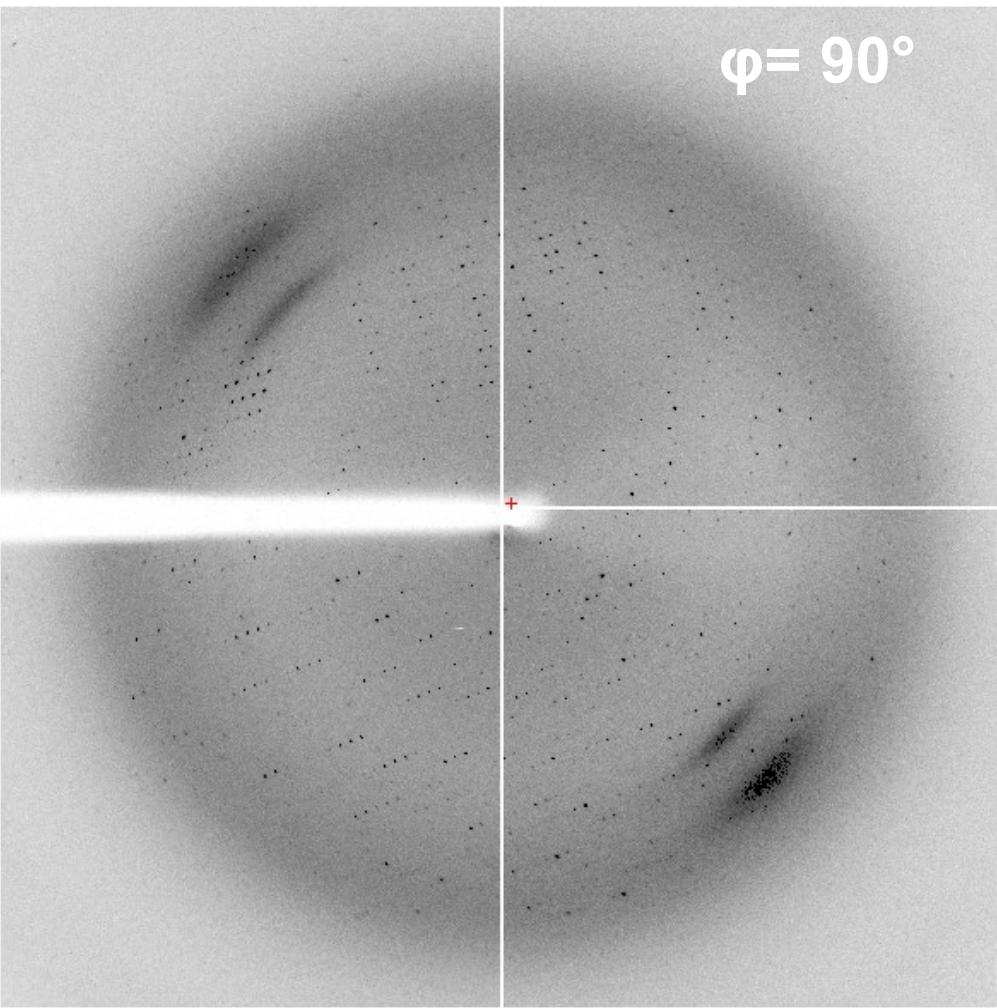
Exposure = 5 s,

Rot. width = 1°,

Resolution=2.8 Å

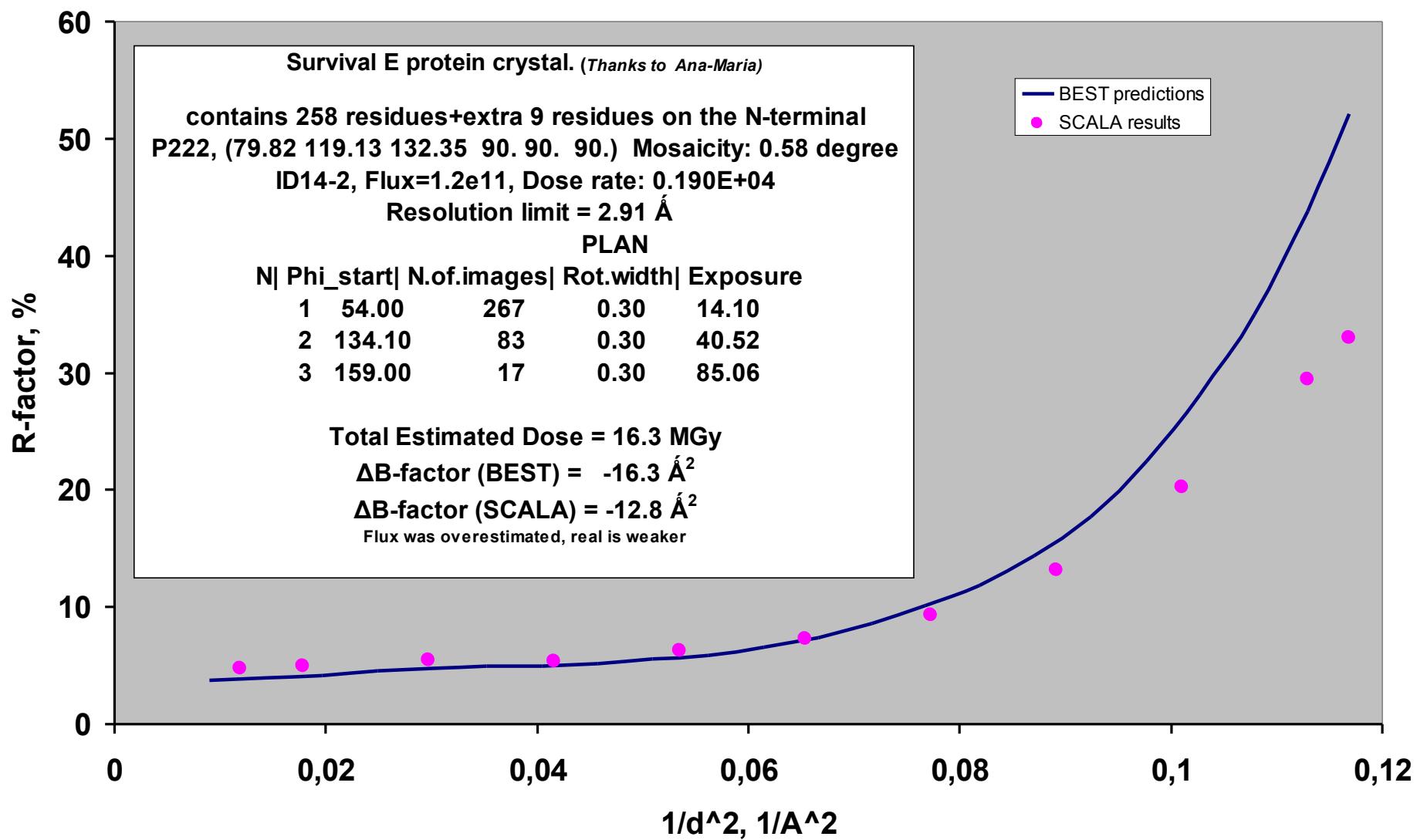


$\phi = 0^\circ$

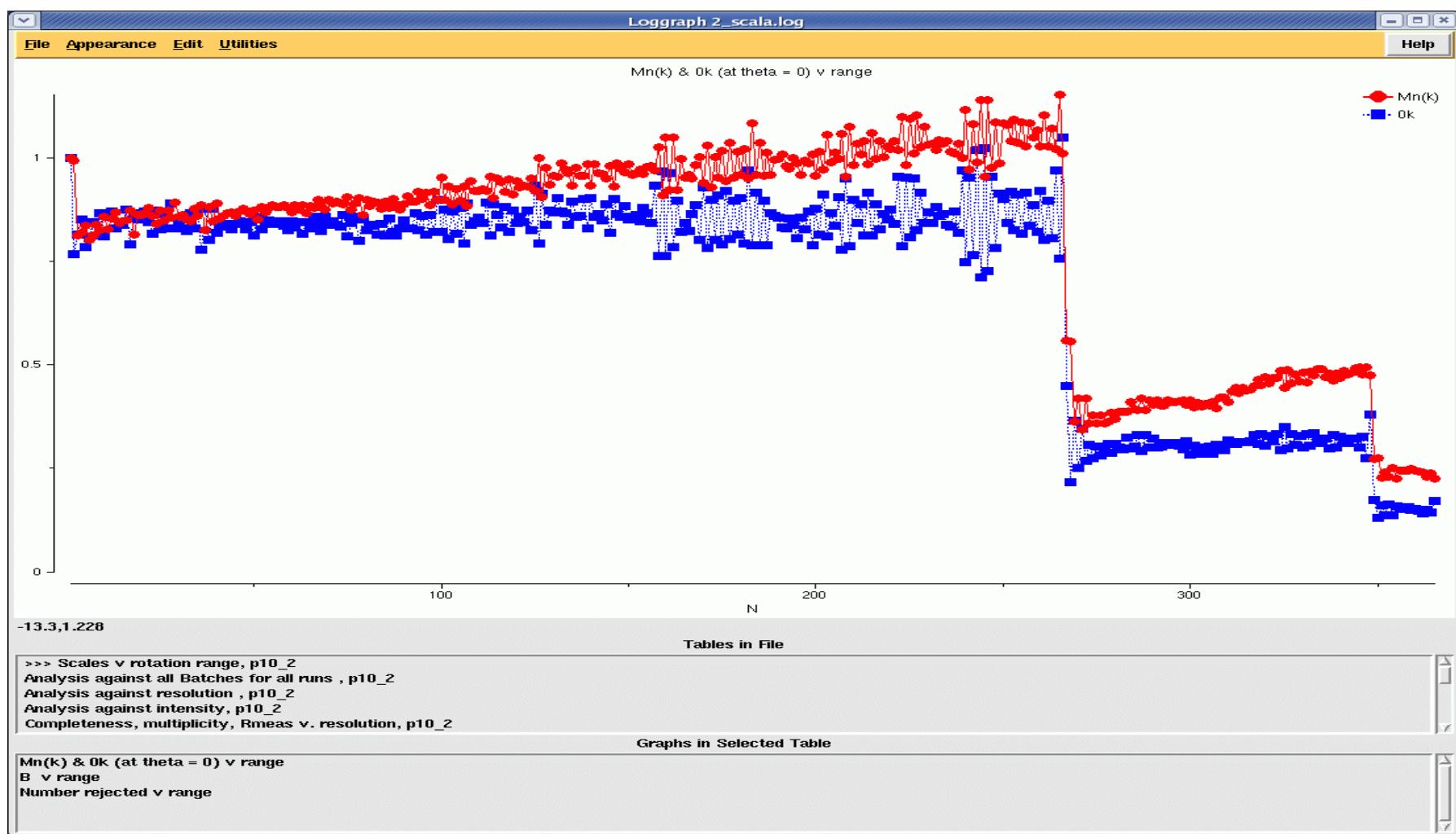


$\phi = 90^\circ$

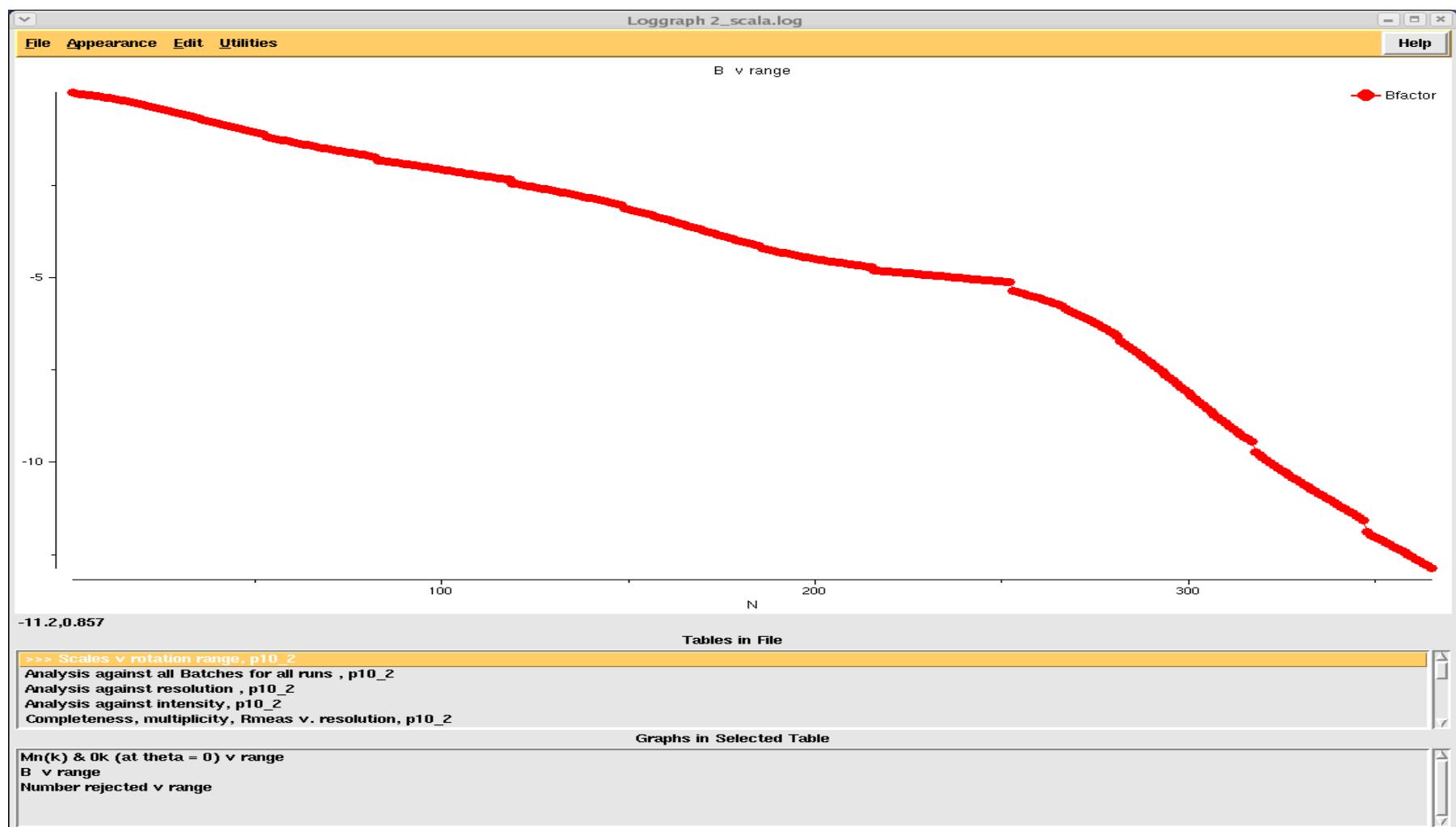
R-factor VS. Resolution



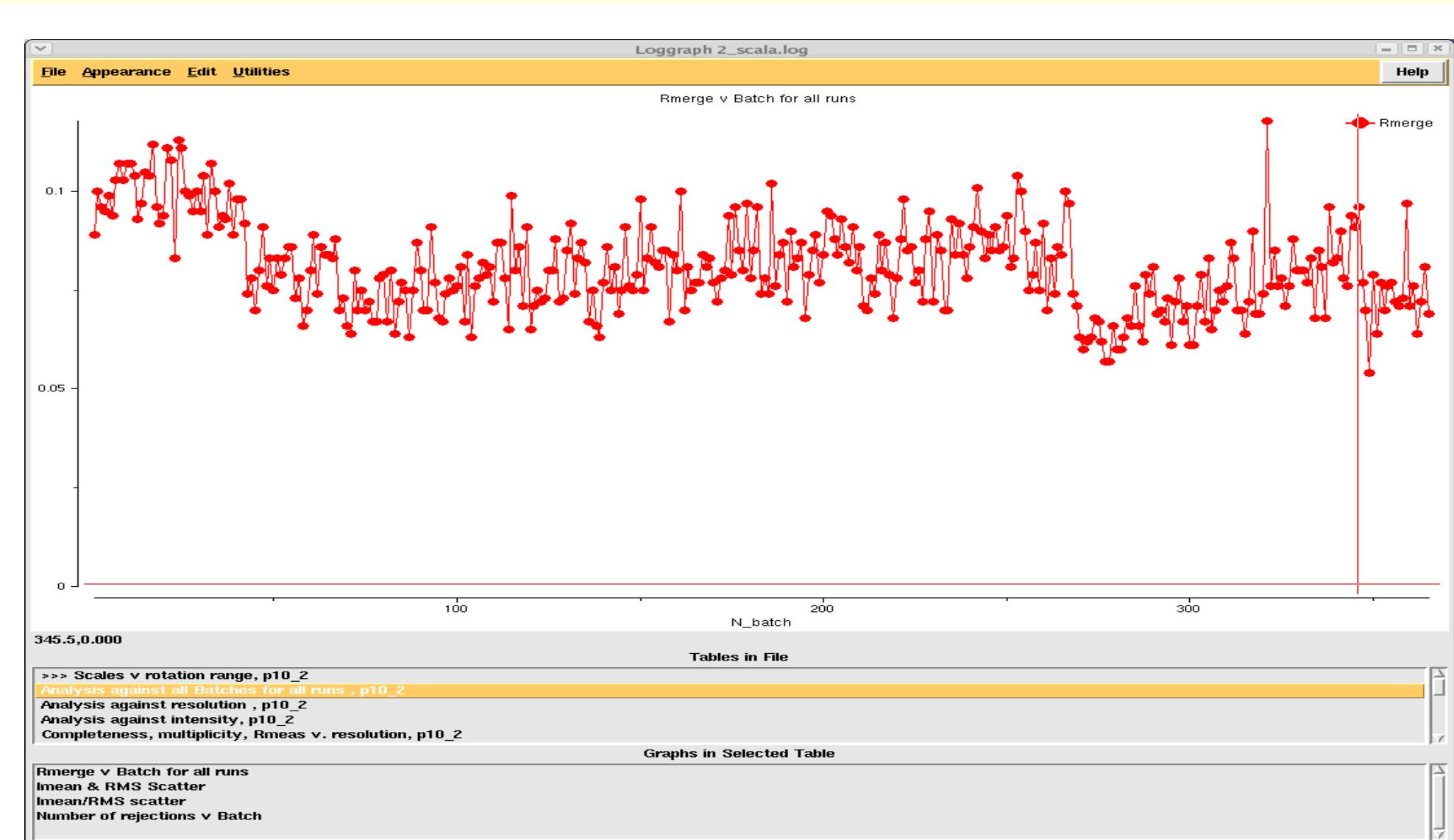
SCALA Output



SCALA Output

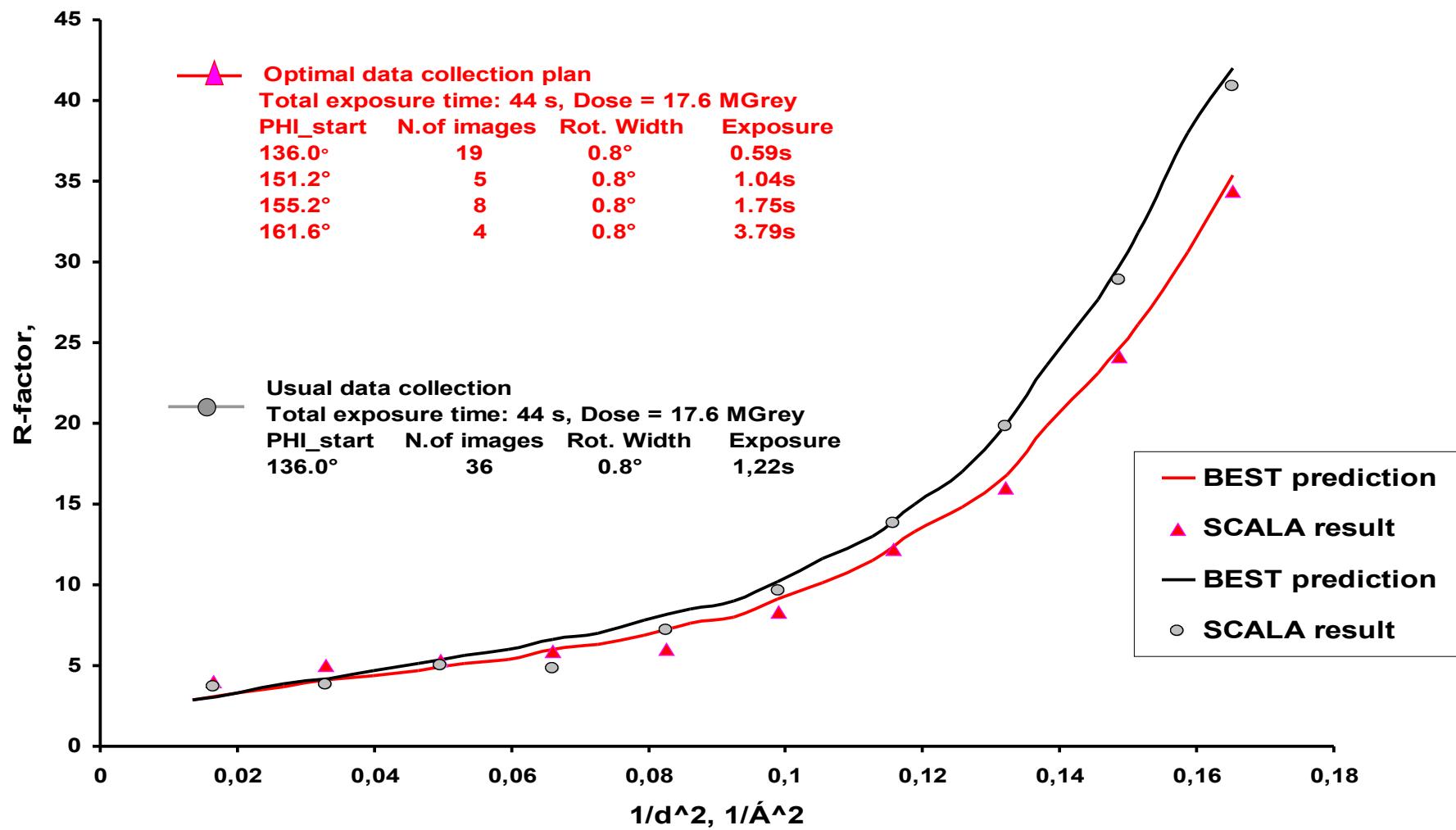


SCALA Output



ID23-1, ESRF

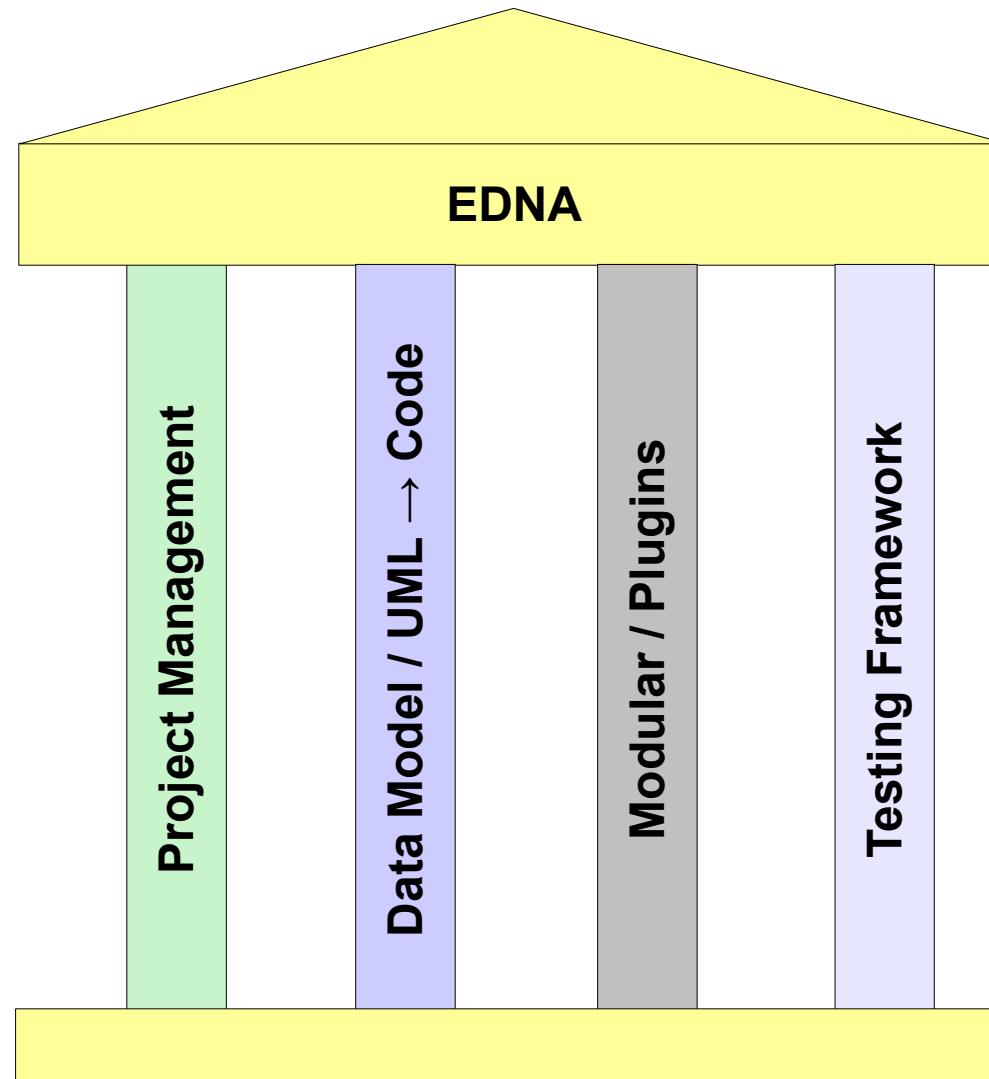
Complex of viral (from tomato bushy stunt virus, TBSV) suppressor of RNA silencing p19 with mismatched siRNA



EDNA is not DNA...

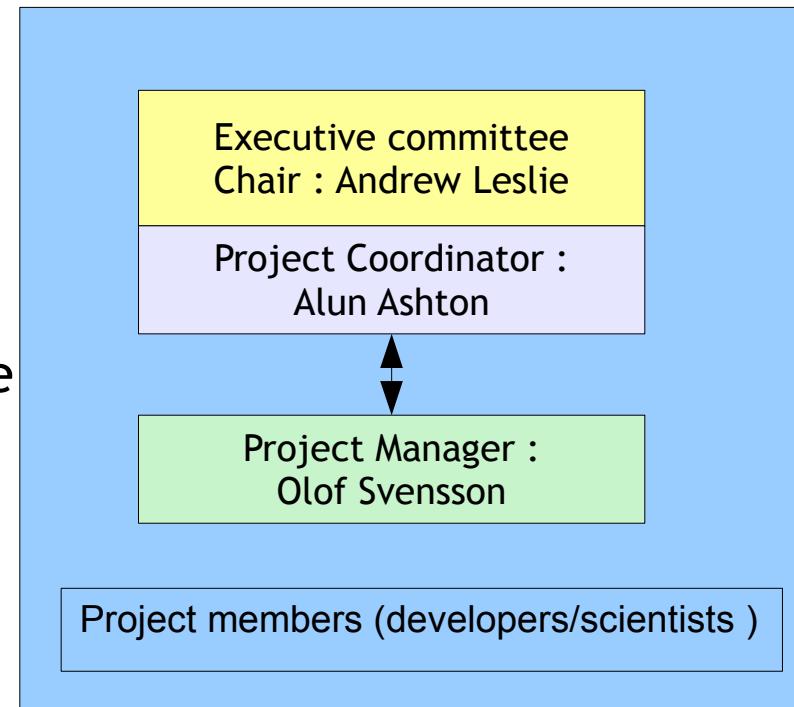
- Same goals (initially)...
 - Automatic MX sample characterisation
 - Online data processing during MX data collection
 - Ranking
- ...however very different implementations
 - EDNA designed to not be specific to MX - separate kernel
 - No shared code base between DNA and EDNA
 - Different project management
 - Different collaborators

The EDNA Project / Framework



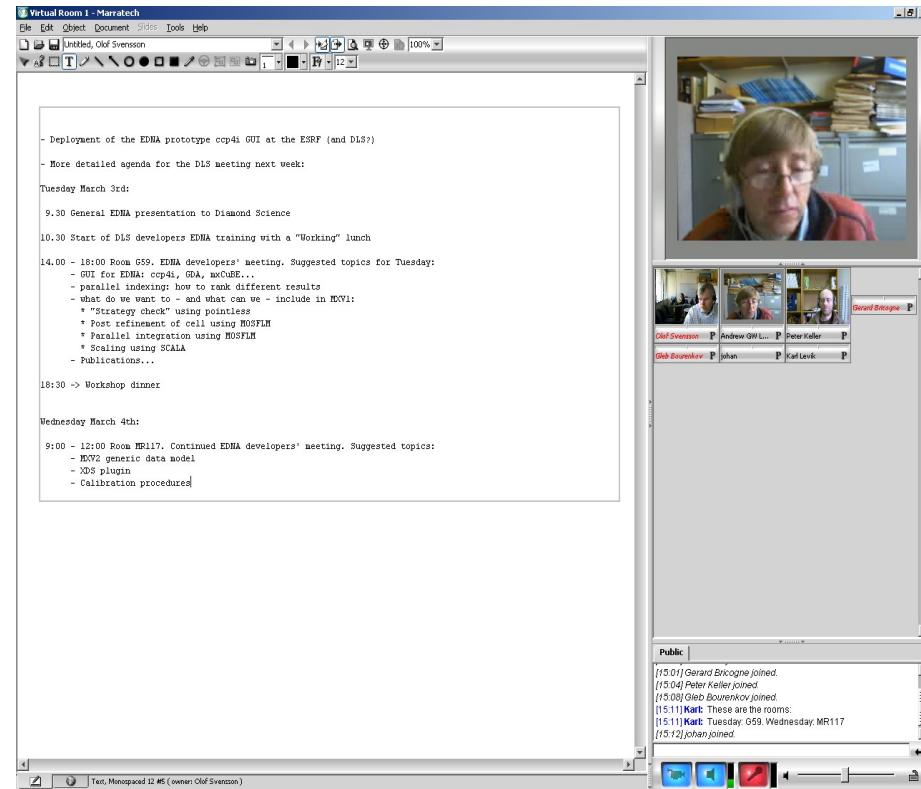
EDNA Project Management (1)

- Executive Committee:
 - Alun Ashton, DLS, UK
 - Gérard Bricogne, Global Phasing, UK
 - Andrew Leslie, MRC LMB, Cambridge, UK
 - Andrew McCarthy, EMBL-Grenoble, France
 - Sean McSweeney, ESRF, Grenoble, France
 - Thomas Schneider, EMBL-Hamburg, Germany
 - Andrew Thompson, Synchrotron Soleil, France
- Other members from:
 - BESSY, Berlin, Germany
 - MAX LAB, Lund, Sweden
 - NSLS, Brookhaven, U.S.
 - SLS, Villigen, Switzerland
 - University of Sydney, Australia
 - University of York, UK



EDNA Project Management (2)

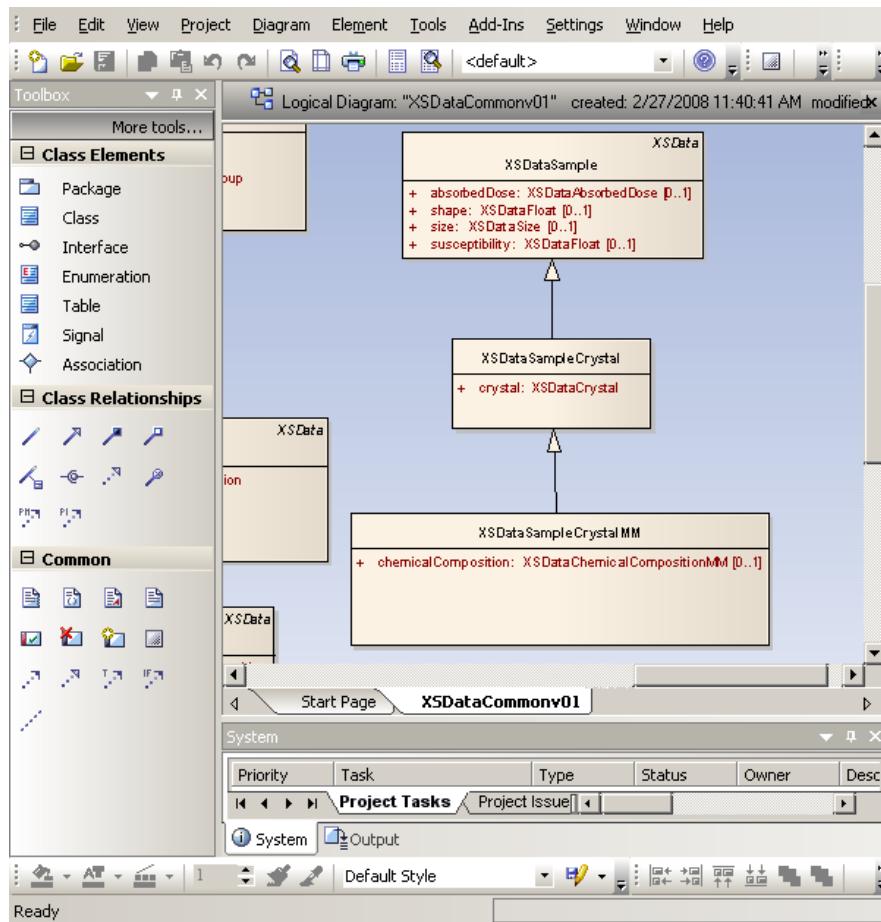
- Project agreement
- Coding conventions
- Code reviews
- Development tools
 - Eclipse
 - Enterprise architect
- Project portal
 - <http://www.edna-site.org>
 - Wiki documentation
 - Bugzilla server
 - Subversion server
 - Discussion forum
- Executive committee
- Video conferences
- Developers' meetings & workshops



Marratech video-conferencing tool

The EDNA Data Model Framework

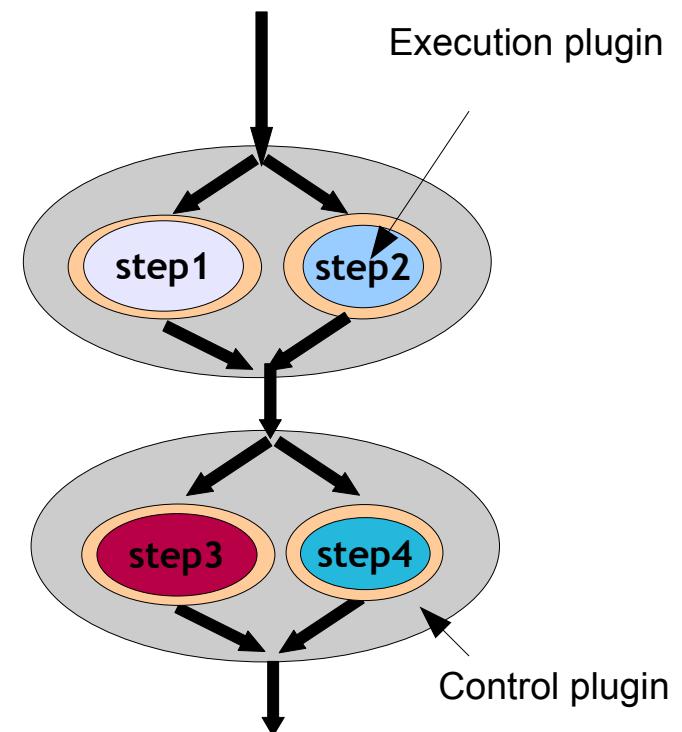
- From UML diagrams to generated code (data binding) :



```
<xs:element name="XSDataSample" type="XSDataSample"/>
<xs:complexType name="XSDataSample">
    <xs:complexContent>
        <xs:extension base="XSData">
            <xs:sequence>
                <xs:element name="absorbedDose" type="XSDataAbsorbedDose"/>
                <xs:element name="shape" type="XSDataFloat"/>
                <xs:element name="size" type="XSDataSize"/>
                <xs:element name="susceptibility" type="XSDataFloat"/>
            </xs:sequence>
        </xs:extension>
    </xs:complexContent>
</xs:complexType>
```

EDNA Modularity : Plugins and their hierarchy

- Plugin base class :
 - Configuration, working directory, etc.
- Execution plugins :
 - Execution of external programs, e.g.
(bash) scripts
- Controller plugins:
 - Control of execution plugins
 - Parallel execution
 - Synchronisation
- EDNA is partly based on AALib,
however AALib is not a part of EDNA
collaboration



EDNA Testing Framework

- The EDNA testing framework consist of three layers :
 - Kernel Unit tests
 - Plugin Unit tests
 - Plugin Execution tests
 - Example: the EDNA Test Suite All result:

```
[SUCCESS] [ 3 ][ EDTTestCasePluginExecuteControlSubWedgeAssemblev10.execute ][1.42272996902]
          [SUCCESS] [ 1 ][ EDTTestCasePluginExecuteControlSubWedgeAssemblev10.testExecute ][1.38881707191]
[SUCCESS] [ 4 ][ EDTTestCasePluginExecuteControlSubWedgeAssemblev10NineImageSubWedge.execute ][2.38712096]
          [SUCCESS] [ 1 ][ EDTTestCasePluginExecuteControlSubWedgeAssemblev10NineImageSubWedge.testExecute ]
```

```
[UnitTest]: #####  
[UnitTest]: EDTTestSuiteAll summary report:  
[UnitTest]: TestSuites: 2  
[UnitTest]: Total TestCases: 48  
[UnitTest]: Total TestCases [SUCCESS]: 48  
[UnitTest]: Total TestCases [FAIL]: 0  
[UnitTest]: [Total TestMethods]: 96  
[UnitTest]: Runtime: 444.1 [s]  
[UnitTest]: Run: 00d:00h:07m:24s:096ms
```

EDNA Collaborators

Alexander Popov^(e)
Alun Ashton^(b)
Andrew Leslie^(h)
Andrew McCarthy^(c)
Andrew Thompson^(k)
Clemens Schulze^(j)
Clemens Vonrhein^(f)
Darren Spruce^(e)
Elspeth Gordon^(e)
Ezequiel Panepucci^(j)
Gérard Bricogne^(f)
Gerrit Langer^(c)
Gleb Bourenkov^(c)
Gordon Leonard^(e)
Harry Powell^(h)
Johan Turkenburg^(m)
Johan Unge^(g)
John Skinner⁽ⁱ⁾

Karl Levik^(b)
Katherine McAuley^(b)
Lucile Roussier^(k)
Marie-Farnçoise Incardona^(e)
Mark Basham^(b)
Meitian Wang^(j)
Michael Hellmig^(a)
Olga Roudenko^(k)
Peter Keller^(f)
Peter Turner^(l)
Pierre Legrand^(k)
Robert Sweet⁽ⁱ⁾
Romeu Pieritz^(e)
Sandor Brockhauser^(c)
Sean McSweeney^(e)
Takashi Tomizaki^(j)
Thomas Schneider^(c)
Uwe Mueller^(a)

(a) BESSY, Berlin, Germany
(b) Diamond Light Source, UK
(c) EMBL, Grenoble, France
(d) EMBL, Hamburg, Germany
(e) ESRF, Grenoble, France
(f) Global Phasing, Cambridge, UK
(g) MAX LAB, Lund, Sweden
(h) MRC LMB, Cambridge, UK
(i) NSLS, Brookhaven, U.S.
(j) SLS, Villigen, Switzerland
(k) Synchrotron Soleil, France
(l) University of Sydney, Australia
(m) University of York, UK

EDNA developers
Executive committee

Future developments

- Release of Mxv1 Characterisation on June 4th:
 - Based on the EDNA prototype
 - Separate kernel
 - Separate MX execution plugins
- MX to be discussed in the EDNA meeting here in Hamburg:
 - Mxv1:
 - Processing of data
 - MXv2:
 - Improved generic data model
 - Implementation of XDS plugins
- Tomov1:
 - EDNA for tomography
- Darc, CCP4 etc.

Why screen and rank?

- On average at the ESRF : 24 data collections per PDB deposition
- Hence it's important to collect data on the best crystals
- Screening : collect reference images from a large set of samples of the same structure and calculate a score for each sample
- Ranking : choose the sample with the highest score
- Without automation : screening is time-consuming and tedious
- Thanks to automation, screening and ranking are fast and easy to use

Screening and Ranking with EDNA

- Screening of an individual sample corresponds to a single DNA / EDNA characterisation
- Different philosophy of EDNA compared with DNA :
 - EDNA has no control of data collection, hence no GUI for screening
 - Screening will be handled by mxCuBE and other beamline control GUIs
- No implementation of ranking yet with EDNA
- Ranking via the ISPyB interface
 - Advantages :
 - Ranking can be made independently of data collection
 - Same interface for DNA / EDNA
 - Disadvantage :
 - No connection (yet) of ranking results to the data collection software (mxCuBE)

Ranking via ISPyB

User: Karl Levik (cnp64921) No warnings.

Role: Admin Proposal: mx0

ISPyB
Information System for Protein crystallography Beamlines

► Data collection ► /mainLayout.jsp

Proposal | **Data collection** | **Samples** | **Shipment** | **Prepare experiment** | **Lab-contacts** | **Feedback** | **Help** | **Logoff**

► View visits
► Search data collections

Selected Visit
Start Date 23-04-2008
BeamLine i03 null
[Back to this visit](#)
[Back to sessions](#)

| Sample Ranking (Rank - Value) | | | | | | | | | | | | | |
|-------------------------------------|-----------------|--------|------------|-------------|------|------|------|------------------|----------------|-----------|-----------------|------------------|----------|
| Select Sample | Image Prefix | Run No | Start time | Space Group | UC a | UC b | UC c | Ranking resol. Å | Exposure times | Mosaicity | Number of spots | Number of images | Total |
| <input checked="" type="checkbox"/> | ref-tls1 | 5 | 22:27:17 | P6122 | 93 | 93 | 129 | #1 1.40 | #1 14.4 | #1 0.24 | #2 3322 | #2 144 | #1 96 % |
| <input checked="" type="checkbox"/> | ref-tls1 | 4 | 16:26:14 | P3 | 93 | 93 | 260 | #3 1.44 | #4 247.8 | #5 0.49 | #1 3682 | #4 169 | #2 66 % |
| <input checked="" type="checkbox"/> | ref-tls1 | 2 | 16:19:22 | P6122 | 94 | 94 | 259 | #3 1.44 | #5 272.9 | #9 0.59 | #5 2919 | #1 128 | #3 64 % |
| <input checked="" type="checkbox"/> | ref-tls1 | 1 | 16:13:55 | P6122 | 94 | 94 | 259 | #2 1.42 | #3 228.2 | #7 0.55 | #3 2944 | #3 146 | #4 63 % |
| <input checked="" type="checkbox"/> | ref-tls1 | 5 | 17:10:57 | P6122 | 94 | 94 | 261 | #5 1.52 | #11 581.2 | #2 0.35 | #8 2792 | #6 200 | #5 61 % |
| <input type="checkbox"/> | ref-tls1 | 4 | 16:11:05 | P6122 | 95 | 95 | 142 | #1 1.40 | #2 143.6 | #11 0.70 | #9 2707 | #5 171 | #6 59 % |
| <input type="checkbox"/> | ref-tls1 | 4 | 22:19:51 | P6122 | 93 | 93 | 130 | #3 1.44 | #14 623.8 | #3 0.40 | #10 2589 | #12 460 | #7 52 % |
| <input type="checkbox"/> | ref-tls1 | 3 | 16:06:08 | P6122 | 94 | 94 | 139 | #4 1.49 | #7 544.8 | #10 0.66 | #6 2902 | #9 301 | #8 51 % |
| <input type="checkbox"/> | ref-tls1 | 2 | 16:09:09 | P3 | 94 | 94 | 259 | #6 1.55 | #10 563.9 | #9 0.59 | #7 2802 | #11 414 | #9 48 % |
| <input type="checkbox"/> | ref-tls1 | 1 | 22:02:35 | P6122 | 93 | 93 | 130 | #7 1.60 | #8 549.7 | #6 0.53 | #14 1987 | #7 263 | #10 48 % |
| <input type="checkbox"/> | ref-tls1 | 2 | 09:17:16 | P6122 | 93 | 93 | 132 | #11 1.79 | #15 631.2 | #8 0.57 | #12 2379 | #8 273 | #11 47 % |
| <input type="checkbox"/> | ref-tls1 | 2 | 22:04:40 | P6122 | 93 | 93 | 129 | #8 1.62 | #16 639.5 | #4 0.46 | #15 1848 | #10 307 | #12 47 % |
| <input type="checkbox"/> | ref-tls1-newbit | 1 | 22:40:20 | P6122 | 93 | 93 | 132 | #5 1.52 | #9 557.8 | #13 0.77 | #4 2936 | #15 692 | #13 45 % |
| <input type="checkbox"/> | ref-tls1 | 3 | 22:10:10 | P6122 | 93 | 93 | 129 | #10 1.69 | #6 539.2 | #8 0.57 | #16 1759 | #7 263 | #14 45 % |
| <input type="checkbox"/> | ref-tls1 | 1 | 15:52:03 | P3 | 94 | 94 | 259 | #12 1.84 | #12 621.6 | #4 0.46 | #13 2303 | #13 480 | #15 44 % |
| <input type="checkbox"/> | ref-tls1 | 1 | 21:52:10 | P6122 | 93 | 93 | 132 | #9 1.67 | #13 623.5 | #12 0.75 | #11 2501 | #14 580 | #16 42 % |
| <input type="checkbox"/> | ref-tls1 | 2 | 22:11:57 | P6122 | 93 | 93 | 132 | #13 | #17 | #14 0.79 | #12 2484 | #16 | #17 20 % |
| <input type="checkbox"/> | ref-tls1 | 3 | 22:15:10 | P6122 | 93 | 93 | 130 | #13 | #17 | #14 0.79 | #12 2465 | #16 | #18 19 % |

Chart Rank Select

Conclusions

- EDNA is a framework for online data analysis
 - Modular
 - Data modelling framework
 - Testing framework
- EDNA is a collaboration
 - Project management
 - Many facilities/institutes participating in the collaboration
- The EDNA MXv1 characterisation application will be released on June 4th
- MXv2 is being developed
- Tomography application are being planned

Acknowledgements

- The DNA team
- The EDNA team
- The ISPyB team
- Sasha Popov for providing me his slides
- EMBL Hamburg for inviting me and hosting this EDNA MX full' meeting
- ...and thank you for your attention!