Biostruct-X Integration: DIALS

Synchrotron and Free-Electron Laser Diffraction Integration

Graeme Winter, David Waterman, Gwyndaf Evans and Alun Ashton

Diamond Light Source, CCP4

July 2012, Harvard Medical School

Overview

- Background - the Biostruct-X project
- What are we doing?
- Why are we doing this?
- How is this different?

Acknowledgements

- Funding: EU FP7 Biostruct-X, Diamond Light Source, CCP4
- Thinking: Andrew Leslie, Phil Evans, Gwyndaf Evans, David Waterman, Gleb Bourenkov, Garib Murshudov
- Collaborating: Nick Sauter, Ralf Grosse-Kunstleve
- Writing: developer community at large

Biostruct-X

- EU framework 7 contract no. 283570
- Main focus of project on trans-national access to facilities for MX, SAXS etc.
- Workpackage 6: data integration and analysis for synchrotron and FEL crystallography

WP6 objectives

- Develop data analysis suite which is forward compatible[1] to:
  - correctly handle pixel array detectors
  - deliver high speed
  - handle high mosaicity, overlapping reflections
  - cope with weak diffraction
- To attempt to unify where possible analysis of synchrotron and FEL data: unified tools for structural biologists, handling of coherence effects.
- Develop tools to manage massive data sets *via* pre-classification[2]

[1] i.e. can evolve with technology
[2] this is more an X-FEL requirement

diamond

WP6 tasks

- 6.1 Development of common frameworks for data management, integration and results inspection (DLS, XFEL, CFEL.)
- 6.2 Development of key algorithm modules for integration program optimised for PAD and challenging cases (DLS, CFEL, XFEL, Dectris, PSI.)
- 6.3 Development of data simulation software and visualisation for raw data and results. Systematic count rate optimisation studies and new hybrid detector technology tests (EMBL-GR, DLS, HZB, CFEL, XFEL, DECTRIS, PSI.)
- 6.4 Development of integration pipeline to handle communication with up and down stream analysis software (DLS.)
- 6.5 Release of integration software to all partner sites for testing (DLS, all.)

diamond

What are we doing? WP 6.1 and 6.2

- Physics-based modelling of experiment
  - Take maximum likelihood approach
  - Work with smart people for algorithm design
  - Emphasise global refinement
  - Enable proper modelling of errors from 1st principles
- Extensible framework
- Build on existing tools
- Speed and smart algorithms

Physics Based Modelling

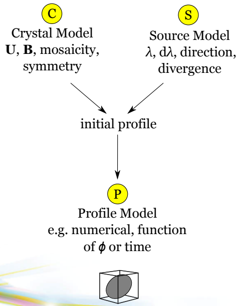
C
Crystal Model
U, **B**, mosaicity,
symmetry

Physics Based Modelling

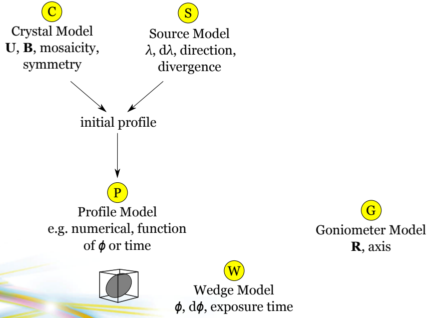
C
Crystal Model
U, **B**, mosaicity, symmetry

S
Source Model
λ, dλ, direction, divergence

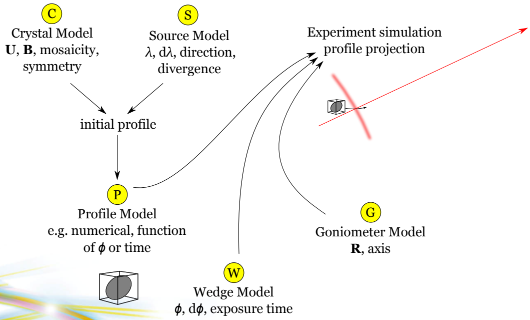
Physics Based Modelling



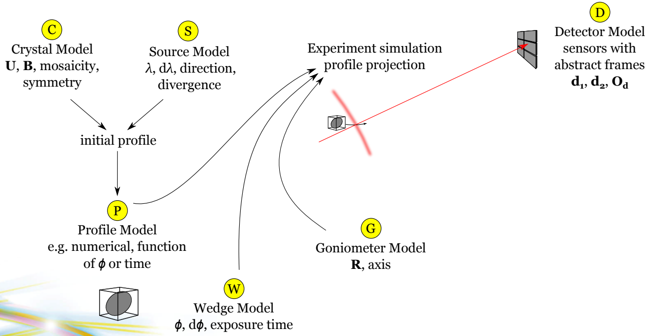
Physics Based Modelling



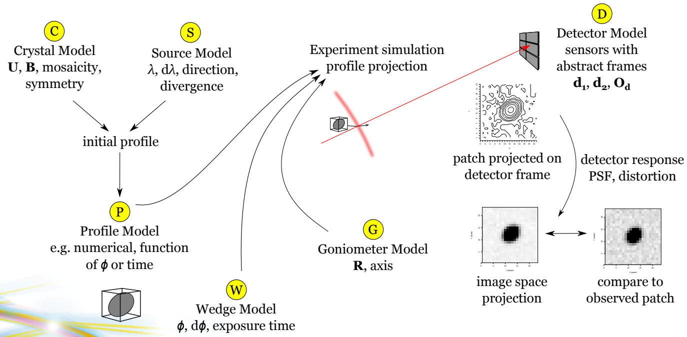
Physics Based Modelling



Physics Based Modelling



Physics Based Modelling



Physics Based Modelling

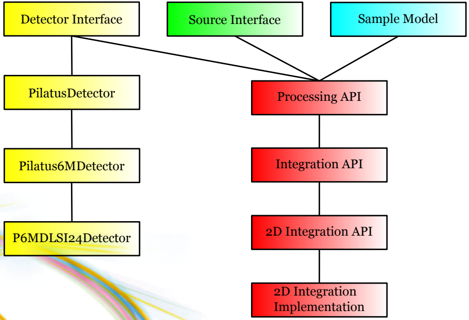
- Illustration had integration on detector surface, on Ewald sphere should also be designed in
- Correctly following equations will be critical for parameter refinement *via* ML - going back from the predicted and observed profiles to model adjustments

Extensible Framework

- Separate out things with proper abstraction layers - should map to real world
- Design system to allow new detector types (and technologies) to be added dynamically c.f. dxtbx
- Also new models for other components e.g. profile modelling
- Options for 2D / 3D / hybrid profiles within same framework
- Synchrotron and free electron laser sources primarily, option for lab sources

Abstraction Layers

An illustration - how the API framework can be organised - which will evolve as the project progresses

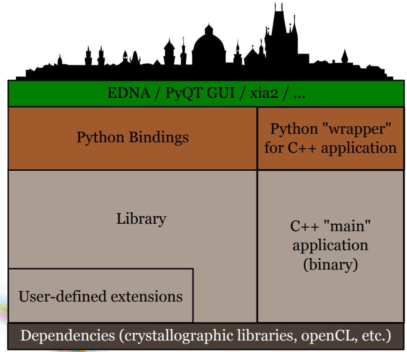


Aside - dxtbx

- Framework to allow *dynamic* detector support extension in *xia2*
- Wrote from the ground up as a toolbox for cctbx - will import into iotbx
- If enough time later will describe this showing code, else can discuss this week

Everything in Library

Not a single user interface - allow operation through GUI, expert system, scripts.



Existing Tools

- Work within and extend CCTBX framework - open source by design, from the start, can piggy-back on their infrastructure e.g. subversion, testing framework etc.
- Collaborating with CCTBX developers, including Nick Sauter (NIH funded project)

Why are we doing this?

- Open source
- Future proof
- Framework for novel algorithm development
- Enable use of modern computing infrastructure

None of these goals are really special - they were desirable in MADNES in the 1980's - 1990's - however the computing hardware was not available at the time to allow this. Now, it is.

Next Steps / ongoing activity

- Define series of use cases (with Nick Sauter):
  - straightforward integration (summation) using masks defined by nearby peaks
  - xds-style "kabsch" transformation
- Defining model components, integrating into use case code
- Starting with well behaved Pilatus data from MX and small molecule crystallography
- Aim at this stage is to determine the nature of the problem, to inform design