Agenda

Meeting title:	CCP4 Working Group 2 meeting	
Date:	Wednesday 28th July 2021	**Time:** 09:30 – 13.00
Location:	zoom	
Circulation:	ccp4wg2@stfc.ac.uk	
Attending:	Pierre Aller, Svetlana Antonyuk, Danny Axford (DA), Arnaud Basle, Ben Bax (BB), John Berrisford, Gerard Bricogne (GB), David Briggs (DB), Nicholas Devenish, Eleanor Dodson (ED), Phil Evans, Maria Fando, Ralf Flaig (RF), Richard Gildea, Deborah Harrus, Mike Hough (MH), Misha Isupov (MI), Ronan Keegan (RK), Eugene Krissinel, Andrew Lebedev (AL), Ed Lowe, Takanori Nakane, Stuart McNicholas (SMN), Peter Moody, Garib Murshudov (GM), Rob Nicholls (RN), Allen Orville, Robin Owen, Randy Read (RR), Alan Roseman, Pietro Roversi (PR), Paul Rowland, Kyle Stevenson, Ivo Tews, José Tincão, Johan Turkenburg, Clemens Vonrhein (CV), Keith Wilson, Graeme Winter (GW), Keitaro Yamashita, Briony Yorke	

Starts 9:30, with 5 min breaks every hour.

1. Approval of minutes from the WG2 meetings 09/06/21
2. Chairs report, introduction and overview (Ivo Tews)
3. Limitations, problems – the user perspective (Peter Moody)
4. Multi-Crystal Methods
    a. Present offers, including DIALS multiplex (Richard Gildea)
    b. GFL multi crystal experience (Clemens Vonrhein)
5. Serial Methods
    a. The use and complexity of XFEL serial data (Allen Orville)
    b. Serial Synchrotron data (Robin Owen)
    c. Indexing stills (Graeme Winter)
    d. Data Quality (Gerard Bricogne)
6. Proposed date of the next meeting 22nd or 29th September
7. AOB

Minutes

1. **Approval of minutes from the online WG2 meeting 9/6/21 (Ivo Tews)**

The minutes from the online WG2 meeting 9/6/21 were approved.

2. **Chairs report (Ivo Tews)**

There have been various meetings to discuss timing and organisation of the meeting. As this is still ongoing, there will be an update in the next WG2 meeting. Discussion concerned the nature of the 'hybrid' meeting, cost, and timing of LTB / What's New / DLS meeting.

During the developers meeting held 7.7-9.7.21, the EDI strategy was explained, as came from discussion in the past WG2 meeting 9.6.21. Also presented were the CCP4 Documentation project that starts 09/21 and Software proposed for inclusion into the suite (see minutes of the last WG2 meetings). NB the path of software was ratified by CCP4 EXEC and is available as a document now (publication on the WebPages).

Brief report on the SWSBC (South Western Structural Biology Consortium) meeting held 14.+15.7.21 online (see http://blogs.exeter.ac.uk/swsbc2021/) that was organised by PostDocs / ECRs

Serial Crystallography – some definitions before start of the discussion (Ivo Tews)

Definition of data types

I. full sweeps
    I. complete: 90 degrees; "crystallographic" multiplicity
    II. or: 180-360-720 degrees, high multiplicity
II. partial sweeps
    I. few degrees, partials and fullys
    II. no internal multiplicity
III. serial data, a.k.a. stills
    I. partial reflections only
    II. no internal multiplicity

DA: consider Synchrotron-Serial, where you may have "time" for a rotation; Robin will ring this up.
GB: in case 2, no internal multiplicity is questionable, for example in virus single shot.
CV: serial (for Synchrotron) can also be defined from the perspective of xtal delivery.
GW: the interesting question is stills *vs.* rotation as even a teeny rotation make a world of a difference.
GW: We ned to consider that XFEL has shot-2-shot variation, while SSX does not.

In scenario 1, questions are which the best datasets to use are, potentially how to combine data, as well as treatment of (highly) anisotropic data, or the dealing of radiation damage. The user will face problems when data processing is not standard, looking at stats, and then decision making.

In scenario 2, the user will have several methods to merge data, but tools would really help the user to do this task, both analytical and decision making.

Scenario 3 is regarded "specialist", and the data may be difficult to process. Various problems exist, even with geometry in the X-Fel experiment, X-Fel detector saturation, scaling and post-refinement, unit cell variation. Analytical tools are needed for best data processing, and there are various gaps to plug. To mention a few: hit detection with improvement on identifying good frames, time-stamps, radiation damage … a single stream can give multiple datasets, scaling and post-refinement? Open for discussion.

3. Limitations, problems – the user perspective (Peter Moody)

Using multiple datasets is not new, this was the method of choice before the advent of cryo: scale within and between subsets is easily done (scale, temperature factor); rejection criteria adjusted manually. This is no longer possible with fine slicing.

Peter went through the accurate measurement of intensities and definition of the measurement box, also in 3D, and compared this with serial data, where this is done by massive redundancy, which sounds wasteful? However, in own experience, XFel data processing, was completely dependent on data processing experts being there. Even with data processing statistics, one has to trust the experts as it is hard to understand what was going on, which Peter exemplified. Documentation is limited. Ideally we would use a CPP4 pipeline that is accessible through the interface.

ITE: On the dose series shown, we integrate over different crystals but don't know the damage inflicted to the individual crystal.

GW: The DIALS West team make great efforts to make XFEL serial user accessible, and there is a new grant funded for the Diamond team to also do this for SSX.

4. Multi-Crystal Methods

a. Present offers, including DIALS multiplex (Richard Gildea)

This chapter is entirely focussed on rotation data, where (1) we have weak data, but everything is considered "the same thing", small wedges; (2) we have RT in situ data, medium wedges; or (3) when increased multiplicity or data quality is needed, e.g. for phasing, medium to full wedges.

Xia2.multiplex can use these scenarios to spit out a dataset, using unit cell clustering, where dials.cosym is a central element that even resolves indexing ambiguity, followed by unit cell refinement and scaling, finished by resolution analysis and space group analysis and isomorphism analysis. Richard explained how this works, going through the steps in detail.

Goal is (near) real-time feedback for more efficient multi-crystal data collection. Richard showed examples from I24 where the question was how many datasets were needed, what radiation damage was, and what data to include. This may also use data form alternative integration programs.

Challenges are combinatorial explosion, and the need to have a reliable metric for the "best" dataset.

The online discussion in the chat additionally covered problems through dehydration of datasets. Also the need to identify the best dataset in a series.

Integration of this capability was seen as necessary.

b. GΦL multi crystal experience (Clemens Vonrhein)

Clemens explained the multi-crystal experience in AutoProc (based on XDS). Looking at this, there are two perspectives: low-dose, high multiplicity data are usually the standard here, and the problem here is rejecting / pruning data; however with small wedges, the problem might be to identify data that fit.

Clemens then went through a couple of scenarios and demonstrated user feedback, using graphical interfaces. These analysis tools are useful to understand what the problems with a given data scenario are. This includes completeness (cantering) and radiation damage. Detection of "poor" data is in three stages, at spot search, after initial integration (P1 usually), and after SG determination. Introduced the "fitness parameter" that combines multiplicity and signal/noise and is (fairly) normalised. Automatic iterative scaling and pruning is being developed – this does not necessarily reflect in R-values but in refinement.

BB: It would be nice to have this for twinning.

5. Serial Methods

Single shot data are serial, and the purest definition is XFEL. However we will consider how small rotations and SSX extend the picture.

a. The use and complexity of XFEL serial data (Allen Orville)

Allen sends a plea for help: do we have enough sample, what is the quality of the data, what is the electron density, and we need this in real time.

Sample delivery: liquid jets are straightforward and can be used for mixing getting a dynamic picture; alternatives are lipidic cubic phases, and concentric mesh injectors. Then we have fixed target approaches which is an efficient way of sample delivery. Acoustic droplet ejection is another way employed.

Allen covered several recent examples. Experiments are typically done in large teams, and the recent solution under the COVID restrictions where only few could travel and people were remote online posed extra challenges. The XFEL hub at X-ray focusses on development of an on-demand sample delivery and X-ray emission spectroscopy. This goes towards access to time resolution, but the question to the audience / CCP4 is: when do we have enough data to have a certain state, in real time?

b. Serial Synchrotron data (Robin Owen)

Robin presented the set-up of I24 and discussed the type of data obtained and the experiments possible. Please consider the information that is currently on the web pages:
https://www.diamond.ac.uk/Instruments/Mx/I24/I24-serial.html

c. Indexing stills (Graeme Winter)

Graeme covered the current state of the art of serial data processing in DIALS. While he consider scaling issues solves, he illustrated that the real problem in serial data processing is indexing the patterns,

d. Data Quality (Gerard Bricogne)

Gerard discussed a specific case where he tickled out problems with processing and treatment of serial data. The talk illustrated that more work will be required, and in particular careful analysis of the merged data.

6. Take note of the date of the next meeting.

Proposed date: proposal 22nd or 29th September 2021, online.

7. AOB

None.