



Agenda

Meeting title:	CCP4 Working Group 2 meeting	
Date:	Wednesday 22 th September 2021	Time: 09:30 – 11.30
Location:	zoom	
Circulation:	ccp4wg2@stfc.ac.uk	
Attending:	Svetlana Antonyuk (SA), Charles Ballard (CB), Arnaud Basle, Ben Bax (BB), John Berrisford, Paul Bond, Lucrezia Catapano, Kevin Cowtan (KC), Eleanor Dodson (ED), Paul Emsley (PE), Maria Fando, Ralf Flaig (RF), Mike Hough (MH), Robbie Joosten, Ronan Keegan (RK), Petr Kolenko, Oleg Kovalevskiy, Eugene Krissinel (EK), Ed Lowe, Airlie McCoy, Karen McIntyre (KM), Stuart McNicholas (SMN), Rob Nicholls (RN), Claudia Milan (CM), Peter Moody, Garib Murshudov (GM), Rob Nicholls (RN), Andy Purkiss, Randy Read (RR), Dan Rigden, Mark Roe, Alan Roseman, Kyle Stevenson, Ivo Tews (IT), Ville Uski, Isabel Uson, Sameer Velankar (SV), Melanie Vollmar (MV), Keith Wilson (KSW), Marcin Wojdyr, Keitaro Yamashita	

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

1. Chairs report (Ivo Tews)
2. ED&I (Jon Agirre, Karen McIntyre)
 - a. CCP4 Meetings: code of conduct
 - b. CCP4 Meetings: code of ethics
 - c. Implementation
3. SW2022 “Current Trends in Macromolecular Model Refinement and Validation” (Svetlana Antonyuk, Melanie Vollmar, Rob Nicholls)
 - a. Program and timeline
 - b. State of invitations and acceptances
 - c. Technical implementation of the hybrid meeting structure
 - d. Plans for the special issue
4. CCP4 and SW2022 offers (Charles Ballard)
 - a. LTB, What’s New (Ville Uski, Ronan Keegan)
 - b. DLS meeting (Ralf Flaig)
5. CCP4 – core group news
 - a. Meetings and workshops (Charles Ballard)
 - b. Updates, v.8.0 (Charles Ballard)
 - c. Core news (Eugene Krissinel)
 - d. Documentation (Maria Fando)
6. AlphaFold – hopes and expectations
 - a. Making >100m structures available through PDBe (Sameer Velankar)
 - b. Sequence search of MrParse against PDB/AFDB (Dan Rigden)
 - c. Using the AlphaFold Database for MR in CCP4i2 (Ronan Keegan)
 - d. Discussion – what do users expect, how does this change crystallography?
7. Take note of the date of the next meeting (proposal 19th January 2022)
8. AOB

Minutes

1. Approval of minutes from the online WG2 meeting 28/7/21 (Ivo Tews)

The minutes from the online WG2 meeting 28/7/21 were approved.

2. Chairs report (Ivo Tews)

Actions from the last meeting – integration of *dials_multiplex* into the suite, t.b.c.

EK: Python 3 dependency, may have to wait for CCP4v8 for full functionality

MV: dials currently on 3.9

Actions from the last meeting – serial processing either at SSX @ Diamond, and as XFEL. There is a need and thus an ongoing effort to integrate this into CCP4. May need further discussion.

EK: test bundles may be in the next release but will require to fork out interfaces.

Actions from the last meeting – path of software into the suite, EXEC approved document on the web page under the About CCP4 > Contributing to CCP4.

Various Alphafold meetings over the summer, see also below.

3. ED&I (Jon Agirre, Karen McIntyre)

a. Implementation (Karen McIntyre)

Extra page on the registration has been added to reflect our efforts and belief. We have taken the Imperial College as a model to develop our strategy.

CCP4 Meetings: code of conduct must be ticked by delegates and speakers.

CCP4 Meetings: code of ethics: there are now a number of queries, all of which has a preference not to say option).

The ED&I group meet regularly, like every 3-4 weeks. KM is a member of the departmental STFC events group that deals with ED&I, and it is clear that CCP4 are further along the ED&I process.

IT: Code of conduct and Code of ethics should be on web pages Many thanks to the team for their efforts.

4. SW2022 “Current Trends in Macromolecular Model Refinement and Validation” (Svetlana Antonyuk, Melanie Vollmar, Rob Nicholls)

The hybrid meeting is planned online and with 220 participants in attendance respecting a 1 m distancing rule. We open registration when all the in-person speakers and LTB speakers are registered.

a. Program and timeline (Rob Nicholls)

Presented the general overview. “What’s New?” is on day 2 in the morning, making the LTB a little shorter. DLS meeting is on day 3 in the morning. There are nine sessions, with session 3 “Prior Information” split in two parts as this is the longest session. Keynotes are day1 and day2 in the evening. All slots are 20 minutes in length, in response to earlier WG2 discussions that for online meetings talks must be short. Introductions and keynotes are 30 minutes. All talk times include 5 minutes for questions, i.e. length of the talk is 15/25 mins.

IT: Enough time to set up desks for registration?

KM: Yes, cleared with the conference centre.

Session structure would be

- 1) Introduction to Macromolecular Model Refinement and Validation
- 2) Practical Considerations
- 3) Prior Information: Restraints and Dictionaries
- 4) New Developments
- 5) Refinement Tools
- 6) Combining Data from Multiple Sources
- 7) Model Building and Interactive Refinement
- 8) Validation and remediation
- 9) Deposition and Bioinformatics

b. State of invitations and acceptances (Rob Nicholls)

Most speakers are accepted at this stage. Some fine tuning going on.

CCP4WG2 welcomed both the programme and the list of speakers.

c. Technical implementation of the hybrid meeting structure (Rob Nicholls)

Equal distribution of speakers online and in-person. Zoom/slack as online media. Chairs are being selected, organisers also to chair. Three chairs per session, chair 1 will introduce speakers, moderate Q&A; chair 2 will do the chat and read out questions; chair 3 is one of the organisers. Timekeeping is important as talks are shorter. Social organisers / hosts / suggestions are sought, and activities should be suited to online and in-person, invitation to go out via BB.

Lucrezia volunteered to help with the social organising.

CM: Not so easy to understand who asks the question. Multiple chairs are good! Make it balanced (gender, across all ages) and do not limit live discussion to known names.

RN: Using questions from the chat makes selection more "objective".

MV: We are using the Webinar style.

CB: Can we ask speakers to answer chat questions in writing.

CM: Agree but note that for Zoom sorting and answering questions is difficult.

MH: Can we consider joint questions at the end of the sessions.

MV: Breaks are generous, so there is some slack.

IT: Can we make recordings available?

KM: Yes, all speakers (except one) have agreed.

RR: Perhaps a good idea for remote speakers to have a pre-recorded talk.

IT: Can we highlight new talent, as we have a good number of young speakers? Can we also make physical attendance available?

GM: Good idea, then you can go and talk to who is there?

KM: Can contact cEvent to add this information.

KC: Chairs need to know that strict timekeeping is important, equality issue!

d. Plans for the special issue (Rob Nicholls)

A large number of speakers have already accepted to publish proceedings (>20). Some exceptions where the papers have just been published.

5. CCP4 and SW2022 offers (Charles Ballard)

Networking events and dinner on day 1 and day 2, all in the larger space.

a. LTB, What's New (Ville Uski, Ronan Keegan)

Ville organises these; there will be a short 30 mins "What's New" and three 45 minutes LTB with gaps in between, running 4-5 in parallel.

IT: How does hybrid work?

RK, RN: two chairs are essential for all workshops.

CM: This even works when speakers are online and the work

b. DLS meeting (Ralf Flaig)

Place on Friday morning is great. Will give an overview and then developments of beamlines, such as XChem and VMX. Also planned are processing pipelines and AlphaFold.

6. CCP4 – core group news

a. Meetings and workshops (Charles Ballard)

The SW2023 will be at Nottingham, booked for 300 delegates. ACA, IUCr David Waterman on ED. IUCr Ville on ccp4online. South American school CCP4 organiser, DLS CCP4 (David), BCA summer school. Workshops are live, not remote or hybrid, as we want the tutors on site, given how intense these are.

GM: Have workshops got a requirement for vaccinations / tests?

CB: Various countries have different regulations.

IT: We need to discuss as CCP4 what policies we want implement to protect speakers and attendants.

CB: It would be difficult to go over the regulations given out by the employer, that is STFC policies. Agreed it would be good to gather what participants expect.

GM: Must not risk that the conference is cancelled because this has not been considered?

b. Updates, v.8.0 (Charles Ballard)

Update 16 coot 0.9.6, MrBump, AlphaFold, monomers, i2 and couled.. update 17 with the last in the 7.1 series. Linux platform is being retired.

Series 8 is based on Python 3.7 with pyside2; however, at the point of release there will be an upgrade to Python 3.9, coinciding with coot for Python 3. For M1 chips, the clocker is QtWeEngine.

c. Core news (Eugene Krissinel)

CCP4cloud with 1813 user, has run 94,218 jobs to date, 170 jobs/day i.e. 80 CPU hours /day, peak 1250 jobs.

CCP4 online runs 600 jobs per month with 100 unique users. There are five standalone installations (mostly industry).

CCP4-EBI grant recruits. Enforced collaboration with EBI also on AlphaFold use, also providing CCP4 tools for using large datasets.

IT: Are online users only power users?

EK: quite dynamic use of these offers.

CB: Ville has statistics from users that show burst use, i.e. about 200 CRANK jobs started at one point.

d. Documentation (Maria Fando)

New project (Southampton/Newcastle/core) that deals with creating a uniform and extendable documentation to guide users.

Three levels of documentation: for developers, use of interfaces, and scientific knowledge.

Starting with CCP4cloud: standard documentation to give the required background and then task specific documentation to give information what is available in a specific task. The documentation project should be an open project where users and developers can add to the cloud.

Articles free of technical details would serve users at the beginners level. There would be tutorials and educational resources. The idea is to establish a framework to develop tutorials on.

ED: I get confused with documentation. Can it all be in Gitlab?

EK: it should all be in Gitlab, but that is just a place to manage it.

KS: Can we access from the CCP4 website.

CM: What is the process of adding documentation, what is asked from the developer to contribute?

EK: This is an open project. The developer goes to Gitlab and adds. The result of this particular project is providing a basis for implementation from a skeleton so that the overall process becomes easier. Please contact us, we will also share this information and contact developers.

7. AlphaFold – hopes and expectations

NEWS: PDB50 meeting 20-22 October is a highlight, please attend.

a. Making >100m structures available through PDBe (Sameer Velankar)

The Cameo team has set up a pipeline www.cameo3d.org/modeling.

Proteome structures will become available UniRef90 cluster (~130 million structures) by end of the year.

3D-beacons.org brings together experimental data with available structures, and in future provide information on complexes.

b. Sequence search of MrParse against PDB/AFDB (Dan Rigden)

AlphaFold is massively better than previous attempts to provide search models for MR. Covariance was needed before, but AlphaFold gets away with much fewer sequences and even designed molecules. Challenges include multi-domain, multi-conformation, and complexes, SNP interpretation or protein folding (interestingly). The main opportunity is MR, where editing/truncation helps (can be automatic). RosettaFold as “inspired by AlphaFold2 before AlphaFold2 was released. Scenarios where running it yourself give an advantage in places, e.g. ensembles and complexes; this is also good for alternative conformations. Download and run locally / cloud, or use Google Colab. Also, there is more knowledge added so running was and when needed can use up-to-date information, making a case for CCP4 cloud installation useful. Developments are: MrParse now reports model, which adds a sequence search as well. Conversion of pLLDT to reasonable B values.

RR: Domain splitting available (Tristan/Tom in CCTBX).

c. Using the AlphaFold Database for MR in CCP4i2 (Ronan Keegan)

Demo on CCP4i2 capability (functionality is also implemented in CCP4cloud). MrBump has now the option to include AlphaFold database under “Search models”; you may turn off the use of the PDB database. The pLLDT cut-offs can also be set, and the B value column can be replaced to replace the pLLDT score to avoid downstream problems with Phaser etc. Integration into MG to select the right truncations, finding much more models; save out for MR.

d. Discussion – what do users expect, how does this change crystallography?

Are users overwhelmed by these changes?

GM: Biology takes central stage, and AlphaFold makes this easier. For developers, what can we make available to experimenters to understand the function of a structure? Challenges are complexes, binding, interaction ... With machine learning, everything is about the data, so complexes are expected to become easier with the

availability of data.

DR: Some amazing complexes are getting predicted already, how about benchmarking?

SV: Being established and is published. The network is not trained to produce diverse structures.

IU: it is important to reflect on the changes and find consensus what and how do we teach structural biology?

SV: We need to train users how to use this information, e.g. the confidence plots.

RK: Could be implemented in MG and MrParse.

IT: Is AlphaFold prediction skewed to energy minimised structures, and complexes, and how can we access diversity?

GM: I think it will be coming, and it will be good if we are part of it.

SV: When there were multiple states in the PDB, then AlphaFold produced different states. What needs to happen is a more systematic approach.

IT: Can AlphaFold structures be trusted and used by the community?

GM: Enzymatic reaction – one fold is not sufficient. Shape yes, potentially even for interactions.

SV: Distance and packing based scores that were used by CASP are part of the model, but an independent metric is now needed. It is also difficult to use the current metric to compare different models.

GM: Validation is a moving target. Validation is now part of the structure determination, which invalidates validation. We need to turn the clock back and leave database focussed approach and come back to physis.

RR: Perfectly valid approach to use a model from AlphaFold, generate hypothesis, and test them.

DR: As long as it is properly validated by experiment I don't see a problem with publishing.

IT: How is CCP4 engaging with this, and responding?

GM: I suggest as one of the next WG2 topics: what do we need structures for?

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

Proposed date: proposal 19th January 2022, online.

9. AOB

None.