Agenda

Meeting title:	CCP4 Working Group 2 meeting	
Date:	Wednesday 9th June 2021	**Time:** 09:00 – 13.00
Location:	zoom	
Circulation:	ccp4wg2@stfc.ac.uk	
Attending:	Jon Agirre, Svetlana Antonyuk, Charles Ballard (CB), Arnaud Basle, Paul Bond, David Briggs (DB), Gabor Bunkoczi, Ivan Campeotto, Lucrezia Catapano, Greg Chojnovski, Kevin Cowtan, Judit Debreczini, Tarik Devron, Eleanor Dodson (ED), Paul Emsley (PE), Phil Evans, Maria Fando, Daren Fearon, Ralf Flaig (RF), Elspeth Garman (EG), Helen Ginn, Mike Hough (MH), Misha Isupov (MI), Robbie Joosten, Ronan Keegan (RK), Nick Keep, Petr Kolenko, Oleg Kovalevskiy, Eugene Krissinel, Andrew Lebedev (AL), Ed Lowe, Airlie McCoy (AMC), Karen McIntyre (KMc), Stuart McNicholas (SMN), Garib Murshudov (GM), Rob Nicholls (RN), Nick Pearce (NP), Anastassis Perrakis (AP), Nikos Pinotsis, Randy Read (RR), Dan Rigden, Mark Roe, Alan Roseman, Pietro Roversi (PR), Roberto Steiner (RS), Kyle Stevenson, Ivo Tews, Johan Turkenburg, Ville Uski, Melanie Vollmar (MV), David Waterman, Keith Wilson, Marcin Wojdyr, Keitaro Yamashita	

Starts 09:00, with 5 min break around 10 am.

1. Approval of minutes from the WG2 meetings 31/03/21
2. Chairs report (Ivo Tews)
3. EDI (Ivo Tews, Jon Agirre, Karen McIntyre)
4. SW2021 digest (Atlanta Cook, Dan Rigden, Anastassis Perrakis) Acta D Special issue
5. SW2022 programme, working title "refinement" (Svetlana Antonyuk, Melanie Vollmar, Rob Nicholls)
    a. Format and session structure
    b. Plans for the special issue
    c. Appointing and contacting speakers
6. CCP4 – core group news (Eugene Krissinel, Charles Ballard)
    a. Updates and news
    b. Meetings and workshops
    c. SWSBC and Northern meeting (Ivo Tews, Ivan Campeotto)

Starts 11:00, with 5 min break around noon.

7. Science
    a. Protein sequence identification and assignment for X-ray and EM (Grzegorz Chojnowski)
    b. Refinement with VagaBond (Helen Ginn)
    c. Shift Field refinement and implementation into i2 (Paul Bond)
    d. Automatic Refinement and Ligand Fitting Workflow in CCP4 Cloud (Oleg Kovalevskiy)
    e. Structure validation and analysis in coot (Paul Emsley)
    f. Integration of Iris into i2: all-in-one representation of different validation metrics for protein models (Jon Agirre)
8. Take note of the date of the next meeting (proposal 29th September 2021)
9. AOB

Minutes

1. **Approval of minutes from the online WG2 meeting 31/3/21 (Ivo Tews)**

The minutes from the online WG2 meeting 31/3/21 were approved online.

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

The SW2022 is presently planned as a hybrid 3-day meeting, with limited attendance y UK residents, only. This format is a departure from previous meeting and would have an extra night on site for attendants, which has financial implications. Therefore, there have been discussions with the core team on cost recovery.

There fallback options are either a two-day hybrid meeting with a one-day online only event or a four-day fully virtual meeting, as in SW2021.

> At this point WG2 had a lively discussion on the nature of the planned hybrid meeting. Rob Nichols, as one of the organisers, explained that this will add two-way interaction to the pure online format. AMC raised the point of insurance if we were to decide on any of the fallback solutions closer to the meeting.

Another discussion since the last WG2 meeting concerned the path of new software into the suite, which was first discussed in WG2 ack in 2018. There is now a document meant to be approved by the CCP4 executive committee. KSW summarised the main consideration as in: dependencies must be identified, the software must run on all platforms (Win/MacOS/Linux), the core team agreeing these are acceptable for distribution, licensing would follow STFC guidelines, and there is a steer to avoid overcrowding of the suite. The path for inclusion would then include presentation in WG2, contact to the core team, and approval by the CCP4 executive committee for inclusion. This process should be published on the website so the process is clear.

> PE asked for clarification on the licence agreement.
> AMC: this is one-sided, can we add the benefits to this document.

3. **EDI (Ivo Tews, Jon Agirre, Karen McIntyre)**

We now have EDI champions Karen McIntyre, Jon Agirre, and Mellanie Vollmar is also interested to join the team. EDI is not simply gender balance, see bioRxiv doi.org/10.1101/2020.04.02.022079. What we next need is Code of Conduct and Code of Ethics implementation.

CCP4 already offer networking and support to ECRs with bursaries, and we have streamed sessions for many years. CCP4 conferences must be further committed to support ECRs. Its about policies and the culture / change.

PE: CCP4 SW is not the same as a workshop, where the pool of talent maybe different.
MV: presented at an XFEL workshop, highly unbalanced both in participants as well as in speakers.
PR: Steven Curry did lead a good discussion on the BB.
MV: There needs to be a lobby for doing this, not just the same pool of people that is convenient, it's about getting asked.
RR: There is still a problem with recruitment.
EG: I am a scientist who happens to be women. Women scientists should not be on the line as women but for the sciences they do. To stand up and put yourself forward requires a lot of confidence. But we need to ensure awareness. There was a good discussion on the postponed South African workshop. The main thing for including women as a speaker must be on own merits, rather than due to gender or other ED&I criteria.

AP: When organising the SW2021, the speaker gender balance came out organically as there was enough choice. However, speaking of ECRs there seems still to be a problem of career transition into permanent post.
RJ: When we organised the SW we were asking group leaders and specifically asked whether younger members of the group could talk about the subject.
KC: Bringing in ECRs addresses some of the bias because the leaky pipeline means that we are selecting against anyone who does not fit the most privileged role.
JA: There are lots of points from the discussion we will take in.
KMc: We used thed IC policy as guide for conferences. STFC and UKRI do a lot in this respect, e.g. flat / accessible rooms, hearing loop, quiet room … but there is more, as this can cover neuro-diversity, sexual orientation, nationality, religion, ethnicity etc. which we need to publicise. We need to make sure that everybody's comfortable - it comes down to having kind and have professionally.

CCP4 ED&I – please contact Karen Mc Intyre and Jon Agirre.

4. **SW2021 digest, Acta D Special issue (Atlanta Cook, Dan Rigden, Anastassis Perrakis)**

There are about 25 articles planned. The status us 1 accepted; 3 by deadline; 6 in may submitted; continuing to chase others. The expectation is that most come which would lead to a final list of minimal 12, but up to 23-25. These submissions seem a fair representation of the Study Weekend.

RR commented that the message about a hard deadline must be commented clearly.

5. **SW2022 programme, working title "refinement" (Svetlana Antonyuk, Melanie Vollmar, Rob Nicholls)**

Changed the working title to “Current Trends in Macromolecular Model Refinement and Validation”.

a. **Format and session structure**

Assuming the three-day hybrid meeting is going forward, the scientific program will be in the PM to arrange for US speakers and audience to participate in the sessions. Within each session, physical/remote speakers may be grouped for tech convenience. Preference is given to shorter talks, allowing for more speakers. The meeting aims to be educational, thus the teaching element is most important when selecting speakers. Young scientists will be considered, or young members in groups, but there should also be exposure given to developers and people that are highly visible from the CCP4bb bulletin board, to put a face to the name. The meeting should provide (historic) background, then cover modern tools and include software from outside the CCP4. Duplication should be avoided, and cohesion across the conference should be achieved. The practical talks would focus on the structure determination procedure / method rather than be structural biology talks.

Preliminary schedule, planned to start Wednesday lunchtime, two scientific sessions each afternoon, lunchtime bytes on Thursday and Friday, dinner/social @ 7.30pm. The mornings would be reserved for “What’s new” and the Diamond Users Meeting, placing them in a prominent position and ideal for UK in-person attendance.

Day 1 – Session 1 Introduction to Refinement
Day 1 – Session 2 Prior information – Restraints and Dictionaries

Day 2 – Session 3 Combining data from multiple sources
Day 2 – Session 4 Refinement tools

Day 3 – Session 5 Model Building and Interactive Refinement
Day 3 – Session 6 Validation, Evaluation and Deposition

Session 1 includes data morphologies (includes twinning, radiation damage, …) that affect refinement and get interesting case studies, suggestions welcome. Session 2 introduces why they are needed, and cover model parametrisation, and then go across resolution scales in refinement, and may include NA. Intention is to cover GRADE, eLBOW and and AceDRG as dictionaries from different packages. Session 3 will point out how to use complementary data, with a focus how this works together, covering Mx, EM, NMX, NMR, and time resolution. Session 4 will focus on refinement packages (other than Refmac, which was covered in sessions 1-3). Session 5 should cover automated and interactive model building, including real space methods and MD driven refinement. Session 6 should then focus on "objective" methods for validation and scientific best practice., and cover deposition and an outlook.

If the meeting were fully remote and extend to four days, we would be able to additionally include more on model/map interpretation and analysis as well as computational techniques, e.g. MD simulations, machine learning.

DB: compare and contrast different suites and methods?
MH: multi-conformer refinement? qFit e.g.
NP: refining ensembles?
DR, MV: refinement and MD in prediction, e.g. alphafold or Rosetta.
NP: 2hrs sessions seem long for remote, also 30 mins breaks seem long, I prefer 15 mins.

b. Plans for the special issue

All speaker with be invited to contribute, and a request to clarification of the intent to submit a paper is made. Articles may be directly related to the subject of the talk but can be delivered on a related topic that is suitable for publication in Acta D. In some cases, a topical review might be acceptable. It is proposed to extend to What's New speakers to contribute, as appropriate.

c. Appointing and contacting speakers

WG2 ratified the proposed programme for the SW2022 and suggests that speakers will be contacted / invited.

6. CCP4 – core group news (Eugene Krissinel, Charles Ballard)

a. Updates and news

CCP4v8 is being developed. There is a considerable change in infrastructure, going to Python 3.7 and Qt5. OSX and Linux bundles are available for testing, developers should adopt these. MS Windows is pending due to compiler change to MSVC to accommodate Qt5 (CCP4i2 already compiles). Alternative MS Windows package for WSL (already available for 7.1 from download pages) that gives good results (including graphics); WSL develops rapidly from WSL-WSL2-WSLG; encouraging, and would safe much effort, and also eliminate Windows specific exclusion, however the focus for now is traditional bundles.

CCP4v7 - there were two updates since last WG2 (Coot, MrBump, check logs). The website and documentation are updated (new style).

The CCP4 Cloud developments may also be useful for running online workshops, with three recent updates since last WG2 (over 30 fixes and new features, e.g. workflows, new integration).

EBI/CCP4 grant awarded (Velankar, Krissinel, BBSRC) – Functional annotations through conformational landscape analysis. Three years PDRAs at EBI and CCP4 starting Feb 2022, please get in touch with EK.

b. Meetings and workshops online

Meetings supported by CCP4:

- Rapidata (Stanford), May 2021

- APS (Argonne, IL), June 2021

Plans for future meetings supported by CCP4 (online unless otherwise stated):

- Developers meeting (Cosener's, UK), July 2021
- Finland (University of Oulu), July 2021
- IUCr 25 (Prague), August 2021 – hybrid!
  EK session chair for online crystallography session and speakers (David and Tarik)
- South America (Montevideo, Uruguay), September 2021
- DLS workshop (RAL/Diamond), Nov/Dec 2021 – attendance?
- SW2022 (Nottingham), Jan 2022 – hybrid
- Israel, Feb 2022 – attendance?

c. SWSBC and Northern meeting (Ivo Tews, Ivan Campeotto)

SWSBC2021 - Going forward as online meeting with ECRs from member universities organising.

Northern meeting – preliminary conversations with PIs across the Midlands (and up North). Responding so far were Leicester, Nottingham, Sheffield, Bradford, Leeds, York, Durham, Newcastle, Edinburgh, Glasgow Dundee. Warwick, Liverpool, Manchester, please contact Ivan if you are interested to join. This is considered as a two-day meeting. Online this year or in-person next year. Not to wait too long would mean end of virtual September, but perhaps to have an in-person meeting means we could have this early next year.

GM: not having a personal meeting but rather virtual would get higher attendance.

7. Science talks

a. Protein Sequence Identification and assignment for X-ray and EM (Grzegorz Chojnowski)

Grzeg explained how to fit unknown proteins to MX and EM with findMySequence, starting form an initial map, using SIMBAD, neuronal networks, and MSA/HMM techniques. The software requires a good model for the backbone. Examples: Lysozyme at different resolution, *Bothrops* venom protein, (published 2021) for MX. In EM the neuronal network is different. Examples: ribosomal proteins with different local resolutions, V-ATPase, GAG protein, ESX-5, demonstrating HMMER is an essential / needed, and the problem with EM maps that leads to fragmented models. Conclusion – findMySequence can identify proteins and can assist model building in difficult cases. Python3 / anaconda, dependencies: cctbx, scipy, h5py, HMMER, pytorch (can be replaced with a C++ module); license: BSD3.

MV: We need machine learning, and pytorch should be included in CCP4.
EK: We should agree on what machine learning we should go for in the suite, maybe discuss in a separate meeting.

WG2 supported putting this forward for inclusion.

b. Refinement with VagaBond (Helen Ginn)

Vagabond is a bond-based refinement, and thus refinement is no longer tied to the geometry. Bond length, and bond angles are rather strict, where torsion angles are allowed to freely vary. The model is defined from a single anchored atom. Example: arginine with an apparent distorted structure, based on 120 overlapping structures all with perfect geometry. Refinement is in real space implemented as a step-search algorithm (gradients / likelihood could be implemented). Future to include electron potential maps / EM. Paper on the basic refinement engine is published.

c. Shift Field Refinement and implementation into i2 (Paul Bond)

Introducing how shift field refinement works, using a gradient map, where the difference map gives a shift in that direction. While the gradient map is sharp and the difference map is noisy, the shifts are therefore also noisy, requiring looking at the difference map over a larger region. The method is therefore good for concerted motions, starting with low resolution first. Complement to jelly body or other refinement modes but shift field refinement is faster than many other methods. To avoid chain breaks, run a limited number of cycles, and apply pseudo regularization. Sheetbend task in CCPi2 can run shift field refinement, and is used by Molrep, Bucaneer and Phaser tasks. Requires LL-BFGS C++ solver.

d. Automatic refinement and ligand fitting Workflow in CCP4 Cloud (Oleg Kovalevskiy)

Automation targets a variety of groups, but mainly novices and inexperienced users, looking at log files and interprete output for them which may otherwise require additional knowledge, and two new tools were implemented. *Verdict* helps with the interpretation of refinement, giving plain text advice, leading to resubmission of a cloned job with modified parameters. *Workflow* will give good initial answers and a starting point for further work which can be followed on by modify / rerun. Developer input needed now to improve Verdict and feedback on useful workflows. The programs can already be accessed.

e. Structure Validation and analysis in Coot (Paul Emlsey)

Several features were presented, including Kekulization of bonds, utilities for handling molecular compounds (fast and interactive), ligand fitting, treatment of conformers and ligand distortion, validation, chemical features for drugs. Coot 0.9.6 announced.

f. Integration of iris into i2: all in one representation of different validation metrics for protein models (Jon Agirre)

IRIS demo shown (from the recently published paper); python module as stand-alone software. All validation criteria plotted concentrical. Implemented in the validation task in CCP4i2, as a completely interactive tool. Because of Python 3.7 and Qt5 dependencies waiting for CCP4v8.

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

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

9. AOB

None.