CCP4 WG1 Meeting 2023 (Hybrid)

5th January

Present: No combined remote/local attendance list recorded.

1. Chair: Opening remarks **David Brown** 10 mins

Dave started by thanking all the CCP4 core team and software contributors for continuing to maintain and develop the CCP4 suite and the user interfaces. He flagged that although, quite correctly, there was a lot of publicity around the developments of Alphafold and Rosettafold there was a lot of hype and the need for experimental crystallographically determined structures was still necessary. Obviously, there is positive impact on the phasing probable with the availability of high-quality models and this was already being leveraged by incorporation into CCP4 pipelines. Dave also thanks the organisers of the Study weekend and the CCP4 support staff. Dave also flagged the significant 8.0 release with strong uptake by users.

Following his opening remarks, Dave Brown stated he was now stepping down as Chair and an election was held for a new Chair. Ivo Tews was nominated by Dave Brown, seconded by Mike Hough and Ivo left the room while the election took place. Ivo Tews was unanimously approved as the new Chair and then asked to re-join the meeting. WG1 thanked Dave Brown for his work as Chair over the last years. Ivo chaired the rest of the meeting.

2. Infrastructure – core team report: **Eugene Krissinel** 10 mins

Eugene summarised the core team members and their roles. **Release 8.0** was achieved on 22 Apr 2022 followed by several updates. Release 8.1 planned for summer 2023.

Release 8.0 highlights Changing Suite's infrastructure to **Python 3** and **Qt5**

- Changing Suite's infrastructure to **Python 3** and **Qt5**
- Automated lines for updates preparation, testing and pushing out
- New software: **ModelcraD, IRIS, PairRef, MrParse, PDB-REDO Tools, Slice-n-Dice**
- **Updated and improved: Coot, Refmac, DIALS, DUI, XIA-2, PHASER, SHELX, Buccaneer, Arp/wArp, BUSTER Interface, CCP4i2, CCP4 Cloud and others**
- **AlphaFold2** functionality integration

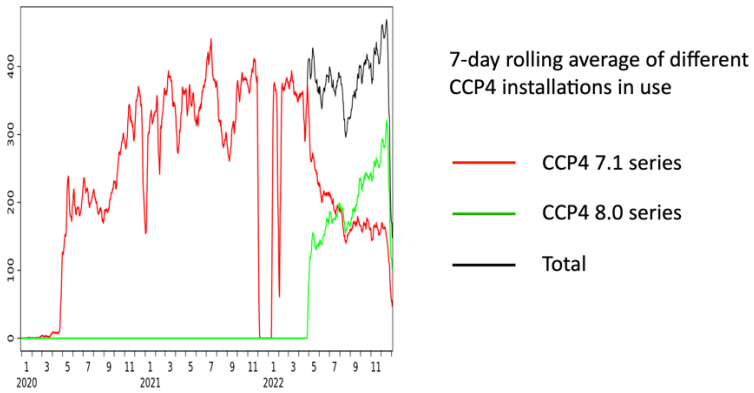
AlphaFold-2 Functionality

- Fully adopted by CCP4 in 2022
- A significant change to structure solution practices
- Capability boost for automatic structure solution
- Seamless support of **AFDB** and (recently) **ESM** databases with **~600M** predicted structures in CCP4 auto-solvers, CCP4i2 and CCP4 Cloud
- AlphaFold/OpenFold software hosted at CCP4-RAL for public use through CCP4 Cloud New **Slice-n-Dice** software to aid phasing from predicted models
- *Future projection for the Suite: more focus on automation, data processing and structure completion*

Software Tests

- Taken care of by Maria Fando after the departure of Oleg Kovalevskyi
- All CCP4 updates are tested automatically before releasing
- Increase number of tests as new software features appear

User statistics



Outreach activities

- Trying to recover from the pandemic situation, back to real-space events
- 2022 CCP4 Schools: APS/Chicago (virtual), SOLEIL (real), AsCA (real), DLS (real)
- Contributions to: Rapidata (Stanford)
- Talks at conferences: ECM-33 (4)
- Webinars: SBGrid
- More plans for Crystallography Schools and contributions in 2023: APS/Chicago, HTCC5/Dubrovnik, MCS2023/Madrid, Crystallography School in Prague
- CCP4 Cloud and associated educational materials are used for university courses (direct feedback from Oxford, Salzburg, Ljubljana)

Main collaborations

Main collaborations

- York/Newcastle: CCP4i2, ModelcraL, Iris, Privateer, Web-Coot
- Liverpool: BioinformaEcs for MR, MrParse, Slice-n-Dice, PISAcov MRC/LMB: Refmac, Coot, Web-Coot
  Global Phasing Ltd: GEMMI, Buster
- Cambridge: Phaser
- Amsterdam/NKI: PDB-REDO
- Leiden: Crank-2
- EMBL-Hamburg: Arp/wArp
- Barcelona: SHELX, Arcimboldo
- Prague: PairRef

Research Projects

- Physical models for Electron Diffraction (WP4 of the CCP4 Grant, BBSRC): *Researcher: Tarik Drevon*
- Functional annotations through Conformational Landscape Analysis (jointly with the EBI, BBSRC): *Researcher: Daniel Garza*
- The group published about 10 papers in 2022, 4 talks given at conferences

Apologies for not giving a complete list

3. Finance: **Charles Ballard** 10 mins

Year	2019-20	2020-21	2021-22	2022-23
Main (£10.5k)	114 (+15)	119 (+9)	132 (+16)	133 (+9)

Subs (£7.55k)	36 (+1)	43	49 (+1)	45
Ac (£3.5k)	5	2	2	3
Total	155	164	183	181

Licenses and income remain strong. There is an expected underspend for the current FY due to increased income which is due to license uptake.

Year	2019-20	2020-21	2021-22	2022-23*
Income	1,681 k	1,748 k	1,681 k	1,900 k
Expenditure	1,491 k	1,640 k	1,654 k	1,820 k
Deferral	1,015 k	1,044 k	1,020 k	940 k or 1,050 k
+/-	-333 k	-79 k	-50 k	-140 k or -30 k

There was/is a delay in filling some posts, which also contributes to the underspend.

However, we are expected to overspend in the next FY.

Year	2022-23*	2023-24	2024-25
Income	1,900 k	1,880 k	1,930 k
Expenditure	1,820 k	2,030 k	1,800 k
Deferral	940 k or 1,050 k	780 k	950 k
+/-	-140 k or -30 k	150 k	40

Study Weekend

- Virtual workshops in 2021 and 2022 have cost about £25k each.
- 2023's smaller hybrid event should cost about £100 k. Registrations in person at 220, with 300 virtual (still open). Number of attendees unclear due to travel disruption.

Workshops

- Sponsored events: SWSBC, Northern meeting
- Full workshops: DLS (Nov 2022), BGU (Feb 2023), APS (Mar 2023)
- Short workshops: ECM (Aug 2022), AsCA (Nov 2022)

4. *CCP4* Current Grant: **Kevin Cowtan** 15 mins

- WP1 (Oxford, Cambridge): Quantifying relationships between diffraction and map quality. Map quality estimation, Fourier optics, X-ray/electron/EM
    - To expand a METRIX database and develop machine learning to support interactive decision making through rapid density map quality assessment. Done.
    - To develop Fourier optics based relationships between diffraction, phase and map quality applicable to scattering experiments including X-ray diffraction, electron diffraction and cryoEM. Done.
    - To develop new tools for refinement against electron diffraction data. In progress.

1. Joosten et al. (2022). Towards Consistency in Geometry Restraints for Carbohydrates in the Pyranose form: Modern Dictionary Generators Reviewed. Current Medicinal Chemistry 29(7), 1193-207.
2. Atanasova et al. (2022). Updated restraint dictionaries for carbohydrates in the pyranose form. Acta Cryst D78, 455-65.

3. Koulouris et al. (2022). Tyrosine 121 moves revealing a druggable pocket that couples catalysis to ATP-binding in serine racemase. Communications Biology 5, 346.
4. Krissinel et al. (2022). CCP4 Cloud for structure determination and project management in macromolecular crystallography. Acta Cryst D78, 1079-89.
5. Yamashita et al. (in submission). GEMMI and Servalcat restrain REFMAC5. Acta Cryst D.

- WP2 (York): Develop shift field refinement methods for Hybrid Structural modelling at low resolution. New techniques for fast combined model rebuilding and refinement.
    - To generalise and optimise shift field methods for wider applicability against multiple data sources. Adapted for EM and refinement of maps rather than models. Done.
    - To implement new resolution independent model building framework in BUCCANEER and NAUTILUS. Increase use of machine learning.. Incorporate structural motif information. In progress.
    - To develop new stereochemical regularisation software library to enable critical speed improvements. Pseudo-regularizer works well, full regularizer for Web Coot. Done.
    - To implement contact predictions developed in WP3 within BUCCANEER and NAUTILUS. Make use of alphafold reliability indicators. To do.

1. P Bond, K Cowtan. ModelCraft: an advanced automated model-building pipeline using Buccaneer. Acta Crystallographica Section D: Structural Biology 78 (9)
2. E Alharbi, P Bond, R Calinescu, K Cowtan. Predicting the performance of automated crystallographic model-building pipelines. Acta Crystallographica Section D: Structural Biology 77 (12)
3. CL Lawson, A Kryshtafovych, PD Adams, PV Afonine, ML Baker, ... Cryo-EM model validation recommendations based on outcomes of the 2019 EMDataResource challenge. Nature methods 18 (2), 156-164
4. K Cowtan, S Metcalfe, P Bond. Shift-field refinement of macromolecular atomic models. Acta Crystallographica Section D: Structural Biology 76 (12)
5. K Cowtan. Structural barriers to scientific progress. Acta Crystallographica Section D: Structural Biology 76 (10)

- WP3 (Liverpool): Implementation and exploitation of contact prediction approaches in crystallographic structure solution and validation. Contact prediction for construct design, model building, validation and refinement.
    - 01/10/19 - 30/09/23. PDRA J. Javier Burgos-Mármol
    - Work continues, in collaboration with Eugene, on extending PISA to include a covariance score to help distinguish biologically relevant interfaces in a crystal structure from lattice contacts
    - CROPS, a spin-off tool developed to help PISACov import sequences and structures from the PDB is fully developed and integrated within PISACov.
    - PISACov provides different scores that are currently being tested against a gold-standard dataset of PISA predictions previously compared against experimental results. Within 2 months we expect to finalise this.
    - Publications. This year should see
    - A PISACov (+CROPS) paper
    - A paper on covariance-based detection of register errors PDB-wide

- New ResearchFish-reportable papers

1. Using deep-learning predictions of inter-residue distances for model validation. Sánchez Rodríguez et al Acta Cryst D78, 1412-1427
2. MrParse: finding homologues in the PDB and the EBI AlphaFold database for molecular replacement and more. Simpkin et al. Acta Cryst D78, 553-559.
3. findMySequence: a neural-network-based approach for identification of unknown proteins in X-ray crystallography and cryo-EM. Chojnowski et al IUCrJ 9,86-97.

- WP4 (Diamond): Physical Model for Electron Diffraction. Modelling electron diffraction, scaling and tools
  - released the T-matrix Simulation Package for ED:
  - HomePage: https://pypi.org/project/pyScatSpheres/
  - Documentation: https://pyscatspheres.readthedocs.io/en/latest/
  - Manuscript: https://www.ccp4.ac.uk/ccp4-ed/documents/articles/pyscat.pdf
  - Python install: pip install pyScatSpheres
  - manuscript on T-matrix formalism for ED accepted for publication in ActaCryst A
  - developed Matplotlib based Gui for simple examples
  - developed online ED simulator: https://www.ccp4.ac.uk/edly/viewer
  - Thickness dependent intensity simulation at arbitrary orientations
  - Analysis of excitation errors
  - Analysis of individual reflections
  - Simulation of rocking curves
  - Comparison with experimental dataset
  - Simulation available with Felix solver
  - Jobs submission to STFC cloud clusters
  - LACBED patterns for dynamical refinement
  - Online comparison with Multislice
  - Conducted ED simulation study of Crambin protein and Biotin and Ireloh compounds, results compared with experimental data from Zendoo, provided by UCLA, Stockholm University and eBIC@DLS collaborators
- Impact (Southampton)
  - CCP4 web page is a success story
  - Driving the ED&I agenda and appointing an ED&I team
  - Completing the CCP4 documentation project building a robust framework for further maintenance and contribution (CCP4 + CCP4i2 + CCP4 Cloud)
  - Organising conferences, workshops and online events in challenging times
  - SW2021 and SW2022 were online with increased participant numbers

- Southwestern and Northern meetings: training and ECR support in the UK landscape
- Establishing CCP4 Cloud as a teaching and learning system; resulted in publication (Acta Cryst D 78, 2022, 1079-89)
- Path of software into the suite through WG2 presentation
- Revised WG2 meeting schedule with more, more inclusive, and shorter meetings, with generally higher attendance numbers (~40)

5. WG2 report: **Ivo Tews** 10 mins

WG2 meetings have a new rota and structure of alternating Science and Housekeeping meetings, 2 hrs, and were held Jan, Feb, Mar, Apr, June, Oct.

WG2 has had several roles / successes, including the organisation of the CCP4 study weekend, overseeing organisation of local meetings, workshops and schools, the CCP4 documentation project, the CCP4 web page, test and release of CCP4cloud, the path of software into the suite via WG2 as published on the web page.

Ivo Tews has been WG2 chair since 2014 and is linked to the role via the impact proposal of the grant. However, in order to smooth transition to the next grant funding phase, IT will step down. Nominations WG2 chair: Jon Agirre and Arnaud Basle (CCP4-Exec to decide at the next meeting January 2023). The role of WG2 chair is linked within the CCP4 structure (reporting in Exec and in WG1).

6. SW 2022 Report. **Ivo Tews** 5 mins

As given and discussed in WG2 (see minutes).

7. SW 2024 Proposals. **Ivo Tews** 10 mins

WG1 need to suggest topics for next Study Weekend. Discussion in WG1. Suggestion to focus the meeting on automation and pipelines. Stress human decision making.

8. Next Grant Application **Dave/Ivo** 15 mins

Several questions were discussed in preparation. Has crystallographic methods development a lesser significance? Hypotheses and answers

Methods development is less important since structure determination is now (mostly) automated.

- The need for crystal structure information is unabated; high quality, atomistic data are required for mechanistic insight: ligand description, chemistry, and interaction.

It's all done, we can now predict structures with AlphaFold2. Complexes soon. Thus we do not require the diversion over the crystal structure any more.

- Even if AI predicts the fold (or even your complex) with confidence, the scientific challenge remains information on structural states and ligand induced change.

Why use a crystal if EM is available on the single molecule level, which additionally can give macromolecular complex and conformation information?

- EM is not necessarily easy or available, and it is not generally applicable; our science case is for a better treatment of electron diffraction data.

The current grant period runs 2019-2024, and a new proposal is needed from 2024. Writing is at an early stage; though ideas are now affirmed. Timeline: discussions mid November @ Exec, this has decided on selection of work packages:

WPx – Liverpool (Rigden)

Hypothesis: Contact / distance predictions from DL methods (AF2, RF2NA, …) can detect model errors

Aim: Implementation beyond single-chain protein validation

Obj1: approaches for protein-protein interfaces, protein NA, and NA models

Obj2: performance studies against the PDB

Obj3: embed in ConKit and in IRIS tool, plugin for WebCoot

Hypothesis: The advance RNA structural biology requires tools for MR and cryo-EM map fitting

Aim: Develop structure solution methodology for RNA using confidence and structural clustering

Obj1: include in Slice'n'Dice (SD) and benchmark

Obj2: AMPLE for RNA models, variance- and confidence-based model editing

Obj3: bespoke recurrent RNA substructure libraries

Methods: DL, AI, covariance/confidence analysis

WPx: ED – RCaH (Krissinel)

Hypothesis: Additional insight into charged states (e.g. ligand binding sites) can be obtained exploiting Electron Diffraction (ED); additionally ED allows investigating nano-crystalline samples.

Aim: Advance computational methods for data processing and structure solution to exploit advantages in ED.

Obj1: statistical methods for data correction and for processing of ED data

Obj2: modelling and treatment of crystal defects

Obj3: methodology for structure refinement in presence of dynamic scattering

Obj4: metrics for ED data, structure solution, and refinement; performance studies

Methods: Simulation of ED data, machine learning, AI

WPx – DLS/DIALS/Soton (lead tbd.)

Hypothesis: Functionally relevant dynamic changes in crystal/s will yield hallmark alterations to subsets of SFs that can be triggered / traced / correlated with dynamic description of macromolecular structure

Aim1: Coordinate model free statistical framework for analysing *change* in data (time and/or state)

Obj1: provide analytical tools / insights into dynamics in near to real time during data collection

Obj2: identify / track / correlate *signature* SF alterations with diffraction (and complementary data)

Aim2: Analysis of molecule driven targets (coordinate / map driven) along a trajectory of change (together with LMB)

Obj1: data selection tools for discrete states

Obj2: trace, model and extract changes in data (trajectory model)

Obj3: refinement against a dynamic data model

Methods: multi-processor SF data analysis; interpretation of low occupancy or multi state; statistical framework to improve quality of data analysis

WPx – LMB (Murshudov)

Hypothesis*: Understanding of small changes and differences in ligands, macromolecules and intermolecular interfaces is key for understanding of functions of molecules*

Aim: Develop mathematical methods, statistical models and software tools for refinement and validation of small changes in molecules

Obj1: Model, refine and validate small differences in molecules using data collected at different states and times. Data may be acquired using X-ray, neutron or electron sources

Obj2: Learn from and build models using the COD and the PDB about ligands, covalent linkages, PTM and apply the derived models for model completion

Methods: Statistical machine learning, Bayesian Statistics, mathematical modelling

The duration of the grant is presently discussed either as phased appointments, or more likely, to cover a shorter time period of 4 years, due to financial constraints. Synergies exist with CCPEM, PDB*redo*, others – should reflect in collaborations? Industry will continue to get support from Core, e.g. Coot, webCoot, ligands, as well as the interfaces GUI2 and *cloud*. Referee? Support Letters?

9. CCP4 Exec: current list of members

Elected		Co-opted	
Dave Brown	Chair 2015 – 23	Eugene Krissinel	Head core team WP PI
Judit Debreczeni	2017-23 Industry Rep	Kevin Cowtan	Grant WP PI
Mike Hough	2020-22	Garib Murshudov	Grant WP PI
Dave Lawson	2021-23	Gwyndaf Evans	Grant WP PI, Diamond Rep
Helen Walden	2022-24	Martyn Winn	STFC Rep
James Murray	2022-24	Dan Rigden	Grant WP PI
		Martin Noble	i2 lead
		Randy Read	
		Keith Wilson	Scientific manager

10. AOB 15:25