CCP4 WG2 Minutes – 25/01/12

Location: Room 3.51 Birkbeck College, London

Present: Phil Evans, Airlie McCoy, Arwen, Roberto Steiner, Nick Kepp, Charles Ballard, Ronan Keegan, Garib Murshudov, Ville Uski, Marcin Wojdyr, Harry Powell, Johan Turkenburg

Chair: Phil Evans

Agenda:

CCP4 Core Group Activities

- 6.3.0 release progress - Charles Ballard
- Improvements to the Windows build – Marcin Wojdyr
- Autobuilds and Nightly builds – Ville Uski

Phaser Updates –Airlie McCoy

Refmac Updates – Garib Murshudov

Aimless/Pointless Updates – Phil Evans

**CCP4 update – Charles Ballard

-new stuff in next ccp4 release:

*mosflm updates (now deals better with pilatus images
*Aimless/scala (major update) see below
*Phaser (major update) see below
*sculptor (search model preparation)
*Ample (ad initio MR pipeline Jaclyn Bibby)
*Dimple (automatic ligand density location pipeline
*Refmac 5.7 (Major update) see below
*ProSMART (structure alignment - external restraint development)
*Zanuda (refinement result analysis, check space group)
*PISA (new update to use SRS databases)
*Arp/warp integration improved. - available for download via ccp4 for academic users

beta of 6.2.999 out very soon - on the dev site.

-CCP4 windows installer created. XP to Windows 7 -Marcin Wojdyr

Overcome compilation efforts.
Windows build now much more automatic (more similar to Mac and linux builds). Additional effort to optimise compiler choices for refmac and phaser

-Nightly builds of CCP4 and downloads – Ville Uski
Now got a buildbot farm for CCP4 - auto compiles and runs tests on software
www.ccp4.ac.uk:8019/waterfall shows status of builds - not externally accessible yet - but will be soon. Builds two versions of linux (ubuntu 11.10 and centos5.6) + MacOSx10.5. Currently thinking about how and when to do the builds. Plan eventually together with Marcin to move towards only rebuilding the stuff stuff that has changed. Tests using Graeme's test frame work.
Also makes a bundle of each build: available from: www.ccp4.ac.uk/dev/nightly
Moving to a new version control (from CVS to Bazaar). Bazaar permits easy branching and merging. Full version history, works on windows. Distributed so different sites can all link to central CCP4.
new downloads page for next release (6.3.0)Test release available on ftp site.
Dowload updating - 1st download a "download" manager and then fire that up to download package.

****Developer's reports**

-Phaser: Airlie
New features in Phaser2.5
modifications address 5 issues with MR in phaser
a) RMS/B factor of model
data and modelling scattering factors all normalised to have a Bfactor of zero coming in. But differences in B-factors between multiple different components is a problem as scattering curve decays differently - causes problems
True RMSD is unknown - what we really want is the rmsd that gives the optimal likelihood through the optimal fractional scattering. Phaser now refines both the bfactors and the variance rms to optimise this. Bfactor refinement is now a default. rms refinement is not a current default as it is quite time-consuming. Improves MR for multiple components - finds all of them, instead of just a subset. Solution list now contains refined Bfactors and rms values.
b) half oligomer problem
Oligomer sat on crystallographic symmetry. I.e. if you're searching with the oligomer MR will fail if the oligomer sits over an axis. Phaser now checks the point group of the model and places any symm ops onto the crystallographic axes and checks if this packs well - assigning a multiplicity value that adjusts for only part of the model being present in the AU. Then adjusts model scattering curves and writes out true au contents. Overlapped chains trimmed in output pdb
c) many copies in the ASU problem
monomer search model, but many copies in the AU. overcomes issue where top 4 hits for tetramer are then each individually placed and tree search triggered and everything gone on ad nauseum. New default method allows multiple copies to be added to the growing tree at each search point. search model can permute. challenge here is what is a real solution? How many of those top solutions do I take? How many of those solutions are real? Currently cuts at 75% difference between top and mean solution. Now: Selects those placements with TFZ> 8. Amalgamation of models must pack (should prune combinations where one or more components have the wrong placement). LLG must also increase for each model placed. Permutation is useful as in phaser the search order is important. So new implementation tests

the effect of using different order of search models to drive solution. Each permutation only allowed to proceed if TFZ>8 after each round. Using this can solve the proteasome! TFZ=X is TFZ from the refinement position. TFZ==X is the TFZ equivalent after amalgamation for each step. 6 hours to solve the proteasome. Not bad.

Airlie commented that a big search of PDB has now firmly established that a TFZ>8 is required to be sure of a correct solution.

d) translational NCS problem

identified by presence of large peak in native patterson. Phasers likelihod functions gove low signal when tNCS is present. Models with low sequence identity that should give a solution without tNCS fail when tNCS is present. Why not just use the tNCS vector and search for two copies separated by tNCS vector. tried this and it doesn't work. tNCS results in systematic modulation of the intensities. can be described by epsilon factors. phaser relies on these being correct as it uses them to weight likelihood functions. Problem is that tNCS never perfect - i.e may not be exactly at the peak and also usually a rotational component. + nonisomorphism causes problems. Use a value tNCS-D as a measure of nonisomorphism. epsilon factors found by refining tNCS-R, -T and -D to maximise likelihood of data. Probability comes from wilson, similar to anisotropy correction.

e) wrong space group problem

previous algorithm - SG identified with highest signal to noise in first translational function when many copies in the AU. This may be wrong is TFZ of the first TF is low- correct SG may only be identified on placement of the final molecule. In practice, all cases seen so far where wrong sg has been identified have tNCS (pseudocentering) - problem is fixed by tNCS code. New algorithm carries the sg with the solution to catch any other sg misidentification cases.

- Refmac - Garib

restraints to external structures now generated by Prosmart. Can identify rigid groups - sequence not used. identifies secondary structures. generates restrains for secondary structures. Robust estimator functions used for restraints. If differences between target and model is very large then their contributions are downweighted. Seems like it allows a flexible way to quickly generate restraints for groups of atoms and then a way to downweight these restraints if the structure is clearly not the same. New button in ccp4i - "use prosmart to generate" -> external restraints to external structure or -> secondary structure restraints. Means if you have a structure you think is similar to yours you can use that provide some initial restraints for your refinement. Needs now a verified and validated fragment library as ideal helix/strand are not good for refinement. Need to test multiple structure restraints - although they're already implemented. Also want to add local ncs averaging (multicrystal as well as single crystal - i.e. 10 structures which should be similar and can restrain each other). Now want to do this for electron density as well.

Tool kits of generation of extra DNA /RNA restraints. base pair restraints should also twists etc.now got a programme that defines base-pair restraints - now generalised to any pair of "monomers". Seem to maintain geometry quite well - but don't penalise structural rearrangements. Same for carbohydrate restraints. For any "monomer" you can define addtional restraints that refmac will read and use.

Sharpening + regularisation now implemented. Improves density considerably. i.e. sausage turning into helix! details clear. example of 4A structure.

Working on a general likelihood function that would work under general assumptions and

account for:
a) experimental uncertainties
b) unmerged data (images may differ from each other, radiation damage)
3) overlaps (twin or otherwise)
4) twins in general form
5) MAD/MIRAS and twin
6) modulation in simple forms (commensurate)
7) extrapolations to higher (and missing data) resolution
Also working on new LMB Murshudov group website. Website will contain bug fixes, dictionary info etc.
Has implemented a kill signal for refmac (i.e. refmac will stop and clean-up after itself if a user defined parameter occurs.)
Q&A: NCS is not applied to Bfactors - just to coordinates

- Phil Evans
--Pointless: all 1.5x versions had a very nasty bug. This is not in the official CCP4 releases
--Aimless: does the same thing as scala - but better. Phil thinks it's ready for prime-time release.
Reads file from pointless and sorts data if needed. Does initial scale estimate, then first round of scaling based on strong reflections. then does first pass of outlier rejection. Optimises profile fitting (weak spots) and summative integration (strong spots) from MSOFLM and makes first estimate of sigmaIs. Then main scaling on relatively strong reflections, chosen on normalised intensity (E^2). Second outlier rejection pass followed by a final optimisation of sigmaIs. Final outlier rejection and produces final statistics. Outputs merged or unmerged data in MTZ or scalepack format (can do all four options from the same job).
Future plan is to combine pointless, aimless and truncate into a single programme. Airlie suggests not truncate. Already a new tab on the CCP4i GUI that wraps these. Aimless now has an autodetect for anomalous signal (can override and it will follow your desires) but if you don't include it Aimless will decide for you. AImless will now read SAINT and XDS files as well as a few other types. For XDS will take INTEGRATE.HKL or ASCII.HKL.
Pointless has an option to take wildcards in the input file name when running from script or command line (note. not when running from GUI). This allows you to merge 100s of files.

SW 2012 Review

Everybody present congratulated Johan and Katherine for the excellent job they did with the 2012 study weekend.

Some minor issues where highlighted such as not enough time for the lunchtime byte sessions and the reduced bar facilities after the main meal.

SW 2013

Three possibilities have been suggested for the topic of the 2013 meeting:

- Data management
- Molecular Replacement

- Protein purification and crystallisation

It was decided that Data management was too limited in its scope and would appeal to too small an audience to be suitable. It was also agreed that Protein purification and crystallisation was outside the scope of what the study weekend should cover and was covered well by plenty of other meetings. As a result, it was agreed the subject for the 2013 meeting should be Molecular Replacement. It will have been six years since the last MR meeting so it's a good time to revisit it in the subject cycle.

This led to the next topic of discussion which was who should be the organisers for the 2013 meeting. Various names were suggested but it was concluded that there should be two organisers Pietro Roversi and Helen Walden with Airlie McCoy as an advisor.

Possible speakers..... (get from Phil)