BUSTER tutorial @ DLS/CCP4 workshop 2020

Introduction

According to ISPyB (https://ispyb.diamond.ac.uk/samples/sid/3177815) we have 5 sweeps of data collected for this Thaumatin crystal:

```
-rw-rw----+ 1 gda2 mx27960_1 139852 Dec  1 10:11 /dls/i04/data/2020/mx27960-1/Test_Thaumatin/Thau1/thauMAD_pk_1_master.h5
-rw-rw----+ 1 gda2 mx27960_1 139852 Dec  1 10:12 /dls/i04/data/2020/mx27960-1/Test_Thaumatin/Thau1/thauMAD_if_1_master.h5
-rw-rw----+ 1 gda2 mx27960_1 139852 Dec  1 10:13 /dls/i04/data/2020/mx27960-1/Test_Thaumatin/Thau1/thauMAD_hrm_1_master.h5
-rw-rw----+ 1 gda2 mx27960_1 139852 Dec  1 10:18 /dls/i04/data/2020/mx27960-1/Test_Thaumatin/Thau1/thauMAD_pk_2_master.h5
-rw-rw----+ 1 gda2 mx27960_1 139852 Dec  1 10:21 /dls/i04/data/2020/mx27960-1/Test_Thaumatin/Thau1/thauMAD_if_2_master.h5
```

(Optional) processing

We could process them all together via

```
process \
  \
  autoPROC_Img2Xds_DamagedPixels="1252,3646 2769,1431 2769,1429 1339,1279 2482,3256" \
  autoPROC_XdsKeyword_UNTRUSTED_QUADRILATERAL="2338 2252 1808 2258 1808 2170 2338 2170" \
  \
  -Id peak1,/dls/i04/data/2020/mx27960-1/Test_Thaumatin/Thau1,thauMAD_pk_1_master.h5,1,3600 \
  -Id infl1,/dls/i04/data/2020/mx27960-1/Test_Thaumatin/Thau1,thauMAD_if_1_master.h5,1,3600 \
  -Id hrem1,/dls/i04/data/2020/mx27960-1/Test_Thaumatin/Thau1,thauMAD_hrm_1_master.h5,1,3600 \
  -Id peak2,/dls/i04/data/2020/mx27960-1/Test_Thaumatin/Thau1,thauMAD_pk_2_master.h5,1,3600 \
  -Id infl2,/dls/i04/data/2020/mx27960-1/Test_Thaumatin/Thau1,thauMAD_if_1_master.h5,1,3600 \
  \
  WavelengthSignificantDigits=4 \
  WavelengthSignificantEnergyDifferenceInEv=0.0 \
  \
  autoPROC_XdsKeyword_DELPHI=8.0 \
  autoPROC_Img2Xds_UseXdsPlugins_DectrisHdf5="durin-plugin" \
  \
  -d autoPROC.01 | tee autoPROC.01.lis
```

There are a few non-default settings here:

- exclude some (not-yet masked) hot pixels
- exclude central part of beamstop holder shadow
- handling of slightly varying wavelength info in meta data (these are values read-out in real-time and not those set by the user) to ensure identical energies are automatically and correctly merged together in the end
- use 80 images per DELPHI block (to make better use of 40 threads in XDS/INTEGRATE)
- ensure we are using the Durin XDS plugin (instead of default Neggia one for master.h5 files)

None of these are crucial (apart from that fix to wavelength meta data) - so one could also just run

```
process \
  -I /dls/i04/data/2020/mx27960-1/Test_Thaumatin/Thau1 \
  WavelengthSignificantDigits=4 \
  WavelengthSignificantEnergyDifferenceInEv=0.0 \
  -d autoPROC.01 | tee autoPROC.01.lis
```

Preparation

For this refinement tutorial, we could also pick up the already processed data directly:

```
/dls/i04/data/2020/mx27960-1/processed/Test_Thaumatin/Thau1/thauMAD_pk_1_/autoPROC/truncate.mtz
```

For more details you can have a look at the summary.html file from autoPROC (/dls/i04/data/2020/mx27960-1/processed/Test_Thaumatin/Thau1/thauMAD_pk_1_/autoPROC/summary.html). In order to perform proper refinement, we also need

1) an adequate model (PDB file), e.g. http://www.rcsb.org/structure/
2) the correct test-set flags (to ensure R-free is not biased at the start)

One way of getting those files is to run a tool from BUSTER:

```
fetch_PDB 6FJ6
```

which will fetch model and reflection mmCIF file from the, perform multiple checks and convert the reflection data into MTZ file (all into a subdirectory called 6FJ6). You can pick those up from DLS disks here (if you don't want to run it yourself):

```
/dls/i03/data/2020/mx27960-4/processing/hcf00057/Projects/Buster/6FJ6/6fj6.pdb
/dls/i03/data/2020/mx27960-4/processing/hcf00057/Projects/Buster/6FJ6/6fj6.mtz
```

The best would be to copy these files first over into an empty directory:

```
cp -p /dls/i04/data/2020/mx27960-1/processed/Test_Thaumatin/Thau1/thauMAD_pk_1_/autoPROC/truncate-unique.mtz .
cp -p /dls/i03/data/2020/mx27960-4/processing/hcf00057/Projects/Buster/6FJ6/6fj6.pdb .
```

Note: the "-p" flag to the copy (CP) command can be very helpful. It preserves the time-stamp of the file ... sometimes a life-saver when trying to figure out what was done several years down the line.

Anyway, what we want to do is to add the test-set flags (and optionally extend, complete or restrict - depending on the different diffraction limits and completeness of the deposited reference data and the newly processed one). An easy way for doing this is to run

```
add_freerflag.sh \
  -m truncate-unique.mtz \
  -f 6fj6.mtz \
  -o autoPROC_peak_1.mtz
```

Refinement

Now we are setup for refinement - the most simple command for running BUSTER would be

```
refine -p 6fj6.pdb -m autoPROC_peak_1.mtz -d BUSTER.01 | tee BUSTER.01.lis
```

As you can see, we are setting

- the PDB model via the -p flag
- the MTZ file with the -m flag
- where output should go: -d defines a to-be-created (sub)directory
- how to save standard output so it doesn't get lost (but still print it to the terminal at the same time)

Some common flags used would be

- -RB (to start with a rigid-body refinement cycle): good for first refinement and/or starting from MR models
- -autoncs (to use NCS restraints): could always be set since neutral if no NCS present - and *always* useful when there is
- -sim_swap_equiv (to sort out NCS relation of symmetrical side-chains like PHE, TYR, ASP, GLU ... often introduced by Coot nomenclature fix)

Remember that BUSTER does several macro-cycles (what we call "big cycles") and that our minimizer takes a certain number of iterations ("small cycles") to reach the function minimum - but that is is very good at actually reaching it. And we want to look at model/maps after convergence since only then we can really draw a conclusion of how to change parametrisation (at the atomic, model level as well the refinement protocol, weighting etc).

Results

After the initial refinement we have a refined model and map available as

```
BUSTER.01/refine.pdb
BUSTER.01/refine.pdb
```

which we can look at via

```
coot BUSTER.01/refine.pdb BUSTER.01/refine.mtz
```

There are some additional useful analysis provided by BUSTER (depending on how the data was processed). Remember that

(1) the data was collected at the Se peak for a selenourea compound, and
(2) by collecting data we generate radiation damage

So it might be interesting to ask those some questions.

Anomalous peaks

If the input MTZ file contained anomalous differences, the final phases from refinement will be used to compute and analyse anomalous Fourier maps. You should see

```
BUSTER.01/diff_fourier.ANO.compare
BUSTER.01/diff_fourier.ANO.pdb
```

Have a look at the first one, e.g. with

```
      head -n 20 BUSTER.01/diff_fourier.ANO.compare
```

which should show something like this:

```
 Peak          Closest atom in BUSTER.01/refine.pdb
 [rms]                                                 Distance (<= 1.0 )
-------------------------------------------------------------------------
  34.07  <=>   O   HOH A 706  (  1.00 19.47)  :        0.26
  32.93  <=>   O   HOH A 734  (  1.00 21.84)  :        0.12
  13.10  <=>   O   HOH A 711  (  1.00 34.81)  :        0.67
  12.22  <=>   O   HOH A 658  (  1.00 27.57)  :        0.49
   9.23  <=>   SG  CYS A   9  (  1.00 15.69)  :        0.09
   8.38  <=>   SG  CYS A  56  (  1.00 17.92)  :        0.14
   8.07  <=>   SG  CYS A 134  (  1.00 16.61)  :        0.09
   7.41  <=>   SD  MET A 112  (  1.00 17.25)  :        0.20
   7.04  <=>   SG  CYS A  77  (  1.00 19.83)  :        0.13
   6.94  <=>   SG  CYS A 204  (  1.00 17.81)  :        0.12
   6.68  <=>   SG  CYS A 158  (  1.00 17.90)  :        0.17
   6.50  <=>   SG  CYS A 145  (  1.00 16.37)  :        0.04
   6.34  <=>   SG  CYS A 126  (  1.00 20.81)  :        0.11
   5.47  <=>   SG  CYS A 149  (  1.00 18.43)  :        0.14
   5.37  <=>   SG  CYS A 121  (  1.00 22.65)  :        0.16
   5.17  <=>   SG  CYS A 193  (  1.00 21.82)  :        0.25
   5.05  <=>   SG  CYS A  71  (  1.00 20.75)  :        0.21
```

This shows all anomalous Fourier peaks that are close to actual model atoms. As you can see, what was modelled as waters (in 6FJ6) has some very strong anomalous peaks: these are probably the selenourea sites (2 strong and 2 weak sites). There are additional (but weaker) peaks at sulfur positions of Cys and Met residues - not unexpected in good data, even if collected at the Se-peak wavelength.

Radiation damage

There is another type of map automatically calculated and analysed: F(early)-F(late) to show potential radiation damage. If data was collected with a large enough rotation range - resulting in high enough multiplicity - autoPROC will have analysed the data to see if a small subset of images (at the beginning of data collection) and another small subset (at the end of data collection) would be complete enough to provide the amplitudes for this kind of map. Please note that this is not a simple re-processing/-scaling of an image subset (which would lead to artefacts due to internal low multiplicity).

We can see a similar analysis to the one above via the following command

```
head -n 20 BUSTER.01/early-late_plus.ISO.compare
```

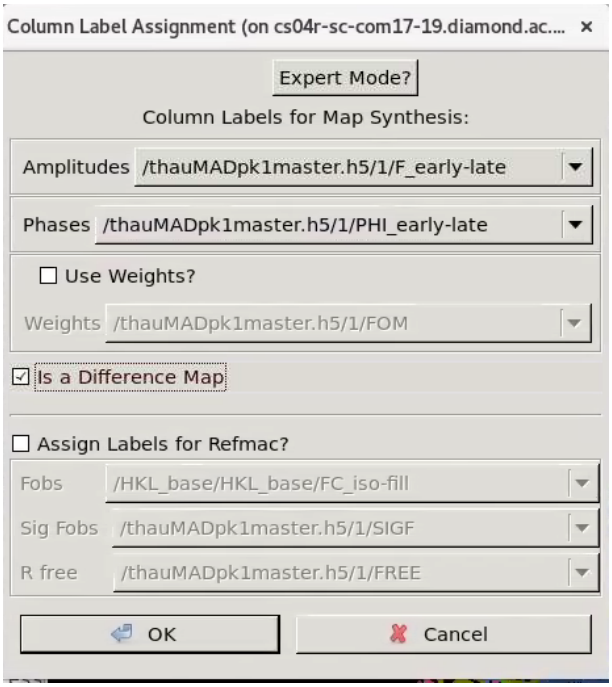
to show us

```
 Peak          Closest atom in BUSTER.01/refine.pdb
 [rms]                                                 Distance (<= 1.0 )
-------------------------------------------------------------------------
   7.00  <=>   SG  CYS A 177  (  1.00 20.63)  :        0.22
   7.00  <=>   SG  CYS A 126  (  1.00 20.81)  :        0.25
   6.46  <=>   SG  CYS A 158  (  1.00 17.89)  :        0.16
   6.21  <=>   SG  CYS A 149  (  1.00 18.42)  :        0.07
   5.32  <=>   OD2 ASP A 147  (  1.00 16.97)  :        0.57
   5.31  <=>   SD  MET A 112  (  1.00 17.25)  :        0.20
   5.21  <=>   SG  CYS A 204  (  1.00 17.78)  :        0.27
   4.92  <=>   SG  CYS A 145  (  1.00 16.38)  :        0.18
   4.87  <=>   SG  CYS A  66  (  1.00 21.71)  :        0.28
   4.83  <=>   SG  CYS A 121  (  1.00 22.65)  :        0.17
   4.82  <=>   O   HOH A 706  (  1.00 19.47)  :        0.53
   4.62  <=>   SG  CYS A 134  (  1.00 16.61)  :        0.19
   4.44  <=>   CA  VAL A 124  (  1.00 17.05)  :        0.85
   4.41  <=>   O   ASP A  70  (  1.00 16.49)  :        0.74
   4.32  <=>   CG  PRO A 141  (  1.00 27.50)  :        0.88
   4.31  <=>   SG  CYS A  71  (  1.00 20.78)  :        0.35
   4.19  <=>   CG  ASP A 186  (  1.00 18.14)  :        0.93
```

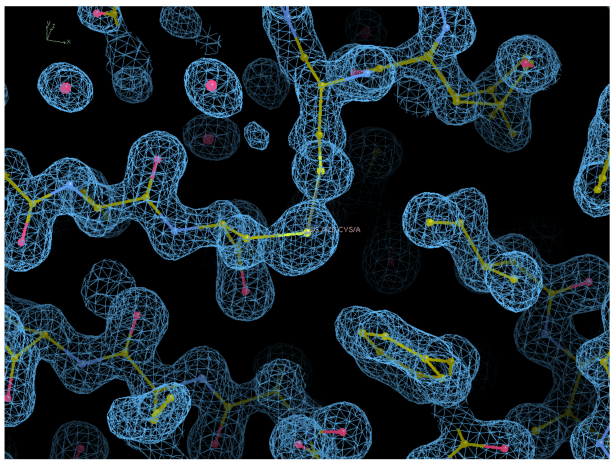
Here you can see (as positive peaks) positions on the model where there is densite at teh beginning of the data collection, but it is missing towards the end, i.e. leading to a positive difference (and peak here). Remember that we collected 5 x 360 degree (3600 images) of data on that crystal - with the last two being a kind of “extra” collection for peak and infl wavelength. What we can see from the above analysis is that already for the first dataset (peak) there is radiation damage at

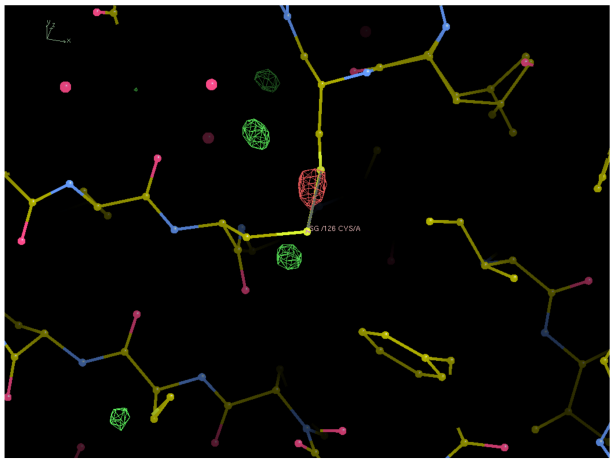
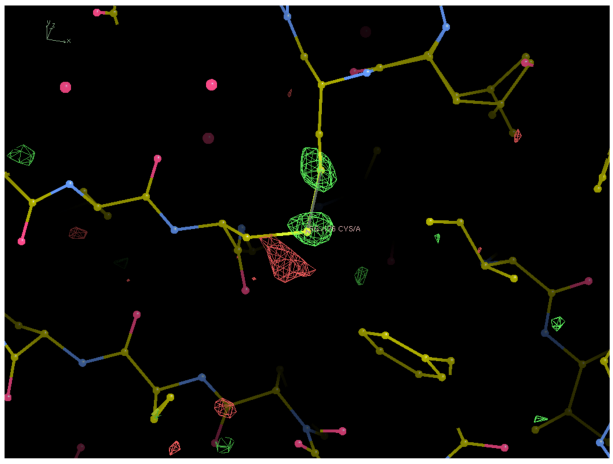
- some of the disulfide bridges (not surprising - they are often the first to break)
- close to an ASP (decarboxylation?)
- close to HOH A706 (which is probably a selenourea site judging from the anomalous analysis above)

You can visualise the F(early)-F(late) map in coot by opening (again) the refine.mtz file and selecting the relevant amplitude/phase columns:



Don’t forget to define this as a difference map. Afterwards you can use e.g. the “Validate => Difference Map Peak Analysis” tool to go through the highest radiation damage peaks (which will be mainly disulfide bridges in this case).

2mFo-DFc map (1.5 rms)	 • Slightly large density around the lower Cys-SG

mFo-DFc map (3.5 rms)	 • Difference map already shows a (partially?) broken disulfide bridge • Is this static disorder (i.e. disulfide bridge sometimes formed and sometimes not in different unit cells of the crystal)?
F(early)-F(late) map (3.5 rms)	 • Radiation-damage detection map shows that the disulfide bridge is present initially and breaks due to exposure.

Continuing refinement

We are now in the normal refinement, map/model analysis and rebuilding cycle ... have a look at our documentation to learn about automatic solvent structure building, TLS, hydrogens handling, occupancy refinement, covalent ligands, ADP refinement, LSSR restraints, ligand detection, restraint generation, ligand fitting, high-throughput ligand screening and much more.

Further information

There is a lot of information on our web pages (documentation and Wiki, examples and tutorials

```
https://www.globalphasing.com/buster/manual/autobuster/buster_reference_card.pdf
https://www.globalphasing.com/buster/wiki/
https://www.globalphasing.com/buster/wiki/index.cgi?DocumentationTopPage
http://grade.globalphasing.org/
```