3CSL

From XDSwiki

HasA/R (PDB id 3CSL (http://www.pdb.org/pdb/explore
/explore.do?structureId=3CSL)) is a complex of a 22-stranded beta-barrel
outer membrane protein (HAsR, 865 residues), its hemophore (HasA, 206
residues), and heme. The structure and its biological implications are described
in "Heme uptake across the outer membrane as revealed by crystal structures
of the receptor-hemophore complex" (Krieg, S., Huché, F., Diederichs, K., Izadi-
Pruneyre, N., Lecroisey, A., Wandersman, C., Delepelaire, P., Welte, W. (2009),
Proc. Nat. Acad. Sci. Vol. 106 pp. 1045-1050.)

3-wl SeMet-MAD data were collected at beamline X06SA of the SLS in
November 2006 on a MarCCD detector. HasA/R crystallizes in spacegroup F222;
cell parameters are a=157Å, b=163Å, c=596Å. There are 2 complexes per ASU.
Data to about 3.0Å could be collected from this crystal, but the anomalous data
are useful to about 5Å only. The ordered part of HasR has residues 112-865 and
harbours 9 SeMet residues. The ordered part of HasA has 173 residues, one of
which is SeMet - but that is mostly disordered.

These MAD data, giving a structure with an average B of 100 Å2, constitute a
project that is challenging for humans, and currently too difficult for automatic
methods of structure solution and model building. The deposited 3CSL
structure was not obtained from these MAD data alone, but the model was
actually refined against slightly better (2.7Å) data collected on a native crystal
at the ESRF.

Contents

XDS data reduction of high-remote, peak and

inflection

The script generate_XDS.INP may be used to get a suitable first XDS.INP file for each of the three wavelengths. Unfortunately the beamline software did not put the correct X and Y position of the direct beam into the header. So you will have to find this yourself, using adxv or XDS-viewer. Or just use:

```
ORGX= 1536 ORGY= 1520
```

The other thing that you might want to try yourself, or just fill in, is

```
VALUE_RANGE_FOR_TRUSTED_DETECTOR_PIXELS=8000. 30000. ! often 8000 is ok
```

instead of the the default (7000. 30000.). This results in a good mask for the beamstop shadow.

It turns out that the spot shapes are actually so irregular that XDS stops after the IDXREF step, with a long warning message. This is because it cannot index (within default error margins) enough reflections (50% is the cutoff). When that occurs, one simply continues with the step after IDXREF:

```
JOBS= DEFPIX INTEGRATE CORRECT
```

Other than that, the three MAD wavelengths can be processed once with default parameters, as written into XDS.INP by generate_XDS.INP. This data reduction therefore proceeds in spacegroup P1, but the correct spacegroup (22) is identified by CORRECT.

Optimization: after this first data reduction pass, I use the "post-refined" geometric parameters, and the correct spacegroup (as given in CORRECT.LP, and written to GXPARM.XDS), for a second pass. Thus I need to

```
mv GXPAM.XDS XPARM.XDS
```

and modify XDS.INP to read

```
JOBS= INTEGRATE CORRECT
```

Afterwards, another xds_par run gives the final intensities. Repeating this optimization sometimes helps.

Peak

360 frames (0.5° oscillation) at the peak wavelength were collected after the high-remote data. They can be downloaded from here (ftp://turn5.biologie.uni-konstanz.de/pub/datasets/3csl-pk.tar) (1.9 Gb). This peak dataset is somewhat difficult to index; if the results are really bad (e.g. distance refining far away from 370 mm) with the default 180 frames, then just try with 90 or 270 frames.

This is an excerpt from CORRECT.LP :

```
REFINED PARAMETERS:  DISTANCE BEAM ORIENTATION CELL AXIS
USING  166758 INDEXED SPOTS
STANDARD DEVIATION OF SPOT    POSITION (PIXELS)    1.80
STANDARD DEVIATION OF SPINDLE POSITION (DEGREES)   0.61
CRYSTAL MOSAICITY (DEGREES)     0.557
DIRECT BEAM COORDINATES (REC. ANGSTROEM)   0.001590 -0.003616  1.021443
DETECTOR COORDINATES (PIXELS) OF DIRECT BEAM   1536.60   1519.03
DETECTOR ORIGIN (PIXELS) AT                    1528.72   1536.95
CRYSTAL TO DETECTOR DISTANCE (mm)       370.85
LAB COORDINATES OF DETECTOR X-AXIS  1.000000  0.000000  0.000000
LAB COORDINATES OF DETECTOR Y-AXIS  0.000000  1.000000  0.000000
LAB COORDINATES OF ROTATION AXIS  0.999995 -0.001537 -0.002692
COORDINATES OF UNIT CELL A-AXIS   -41.622  -128.332    80.636
COORDINATES OF UNIT CELL B-AXIS    45.889    72.744   139.459
COORDINATES OF UNIT CELL C-AXIS  -551.193   220.474    66.370
REC. CELL PARAMETERS   0.006362  0.006103  0.001674  90.000  90.000  90.000
UNIT CELL PARAMETERS    157.174   163.848   597.351  90.000  90.000  90.000
E.S.D. OF CELL PARAMETERS  7.9E-01 8.7E-01 3.0E+00 0.0E+00 0.0E+00 0.0E+00
SPACE GROUP NUMBER     22

...


   a        b       ISa
7.764E+00  7.144E-04   13.43

...

     NOTE:      Friedel pairs are treated as different reflections.

SUBSET OF INTENSITY DATA WITH SIGNAL/NOISE >= -3.0 AS FUNCTION OF RESOLUTION
RESOLUTION     NUMBER OF REFLECTIONS    COMPLETENESS R-FACTOR  R-FACTOR COMPARED I/SIGMA   R-meas  Rmrgd
  LIMIT     OBSERVED  UNIQUE  POSSIBLE    OF DATA   observed  expected

   8.29      26331    7014     7090       98.9%      5.3%      5.5%    26314   20.22    6.3%    4.
   5.90      48060   12542    12555       99.9%      8.4%      8.4%    48060   13.36    9.8%    8.
   4.83      61534   16074    16144       99.6%     11.0%     10.6%    61534   10.69   12.9%   13.
   4.19      71665   18994    19085       99.5%     12.5%     11.8%    71658    9.61   14.6%   16.
   3.75      77668   21598    21677       99.6%     19.5%     19.3%    77668    6.33   23.0%   27.
   3.42      78594   23767    23865       99.6%     28.3%     29.5%    78582    4.14   34.0%   45.
   3.17      64135   24351    26036       93.5%     42.7%     46.4%    60830    2.18   52.9%   78.
   2.97      40861   20207    27920       72.4%     63.8%     72.3%    35172    1.18   83.3%  118.
   2.80      23238   15074    29715       50.7%     89.5%    104.9%    15359    0.63  122.8%  175.
  total     492086  159621   184087       86.7%     14.8%     15.1%   475177    6.17   17.8%   30.


NUMBER OF REFLECTIONS IN SELECTED SUBSET OF IMAGES   501334
NUMBER OF REJECTED MISFITS                             8429
NUMBER OF SYSTEMATIC ABSENT REFLECTIONS                  0
NUMBER OF ACCEPTED OBSERVATIONS                      492905
NUMBER OF UNIQUE ACCEPTED REFLECTIONS                159845
```

The verdict is clear: high mosaicity (> 0.5°) bad ISa, anomalous correlation > 30% only to about 5 A. The reason becomes clear if we load FRAME.cbf into XDS-Viewer, and zoom in:
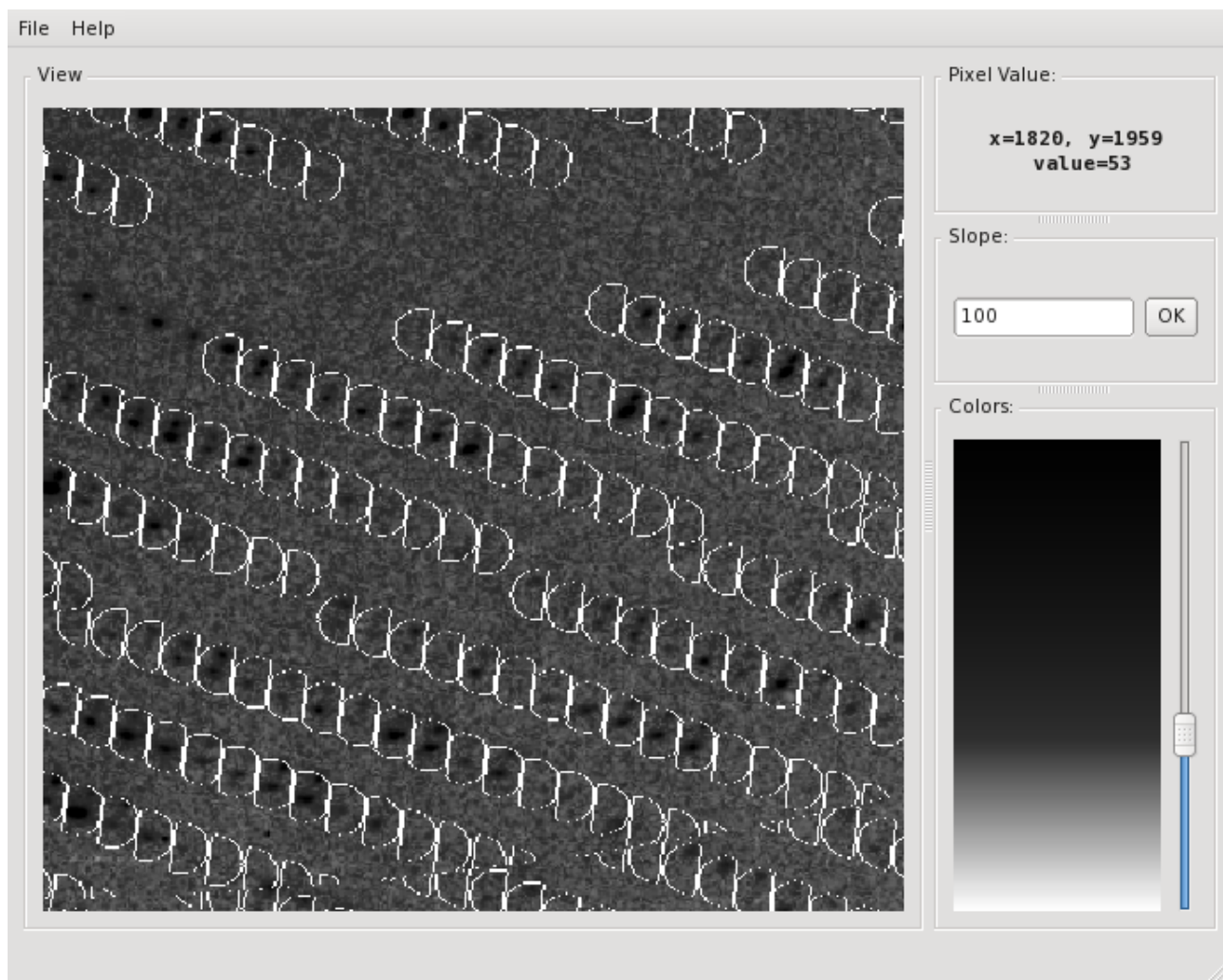
It is clear that these split reflections provide bad data. Fortunately it seems that in other areas of the detector, the reflections look better.

We can use the "scalefactors" jiffy to investigate the scale factor, and the estimates for mosaicity and beam divergence of each frame:
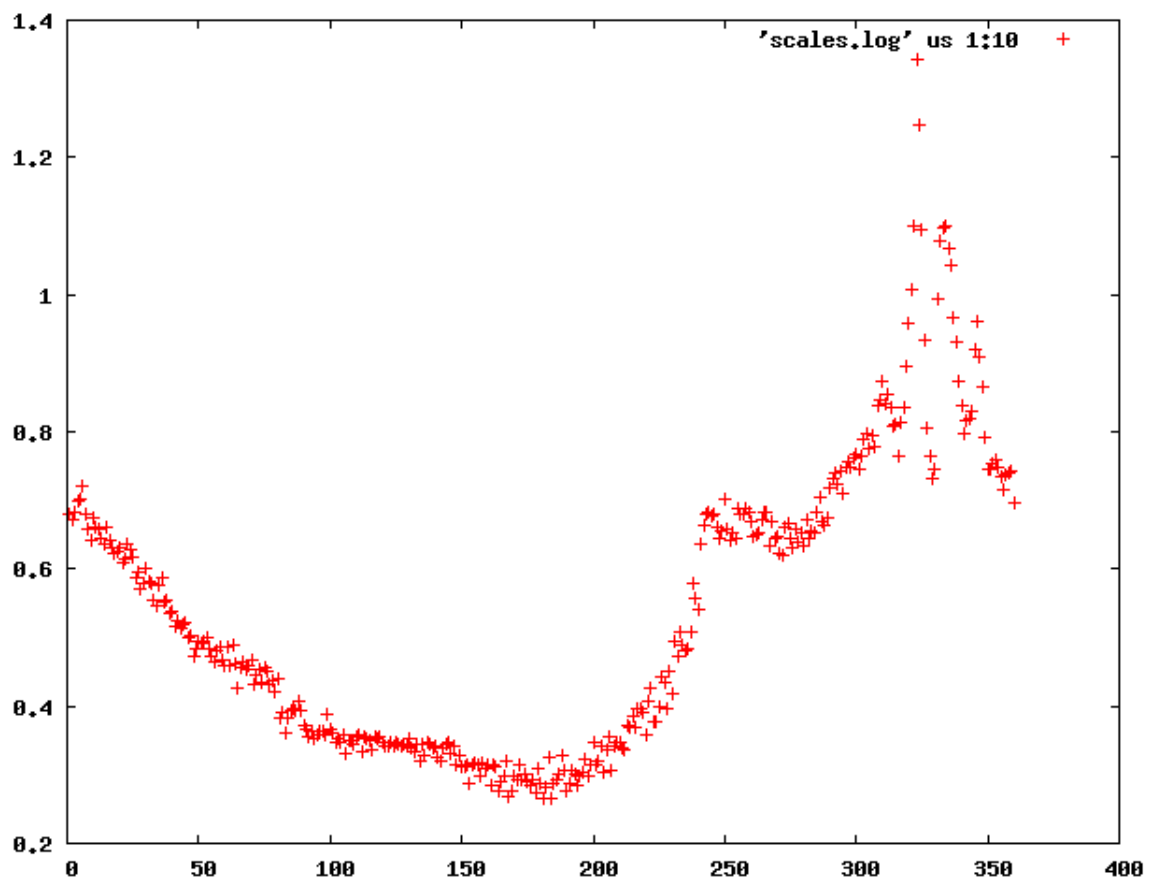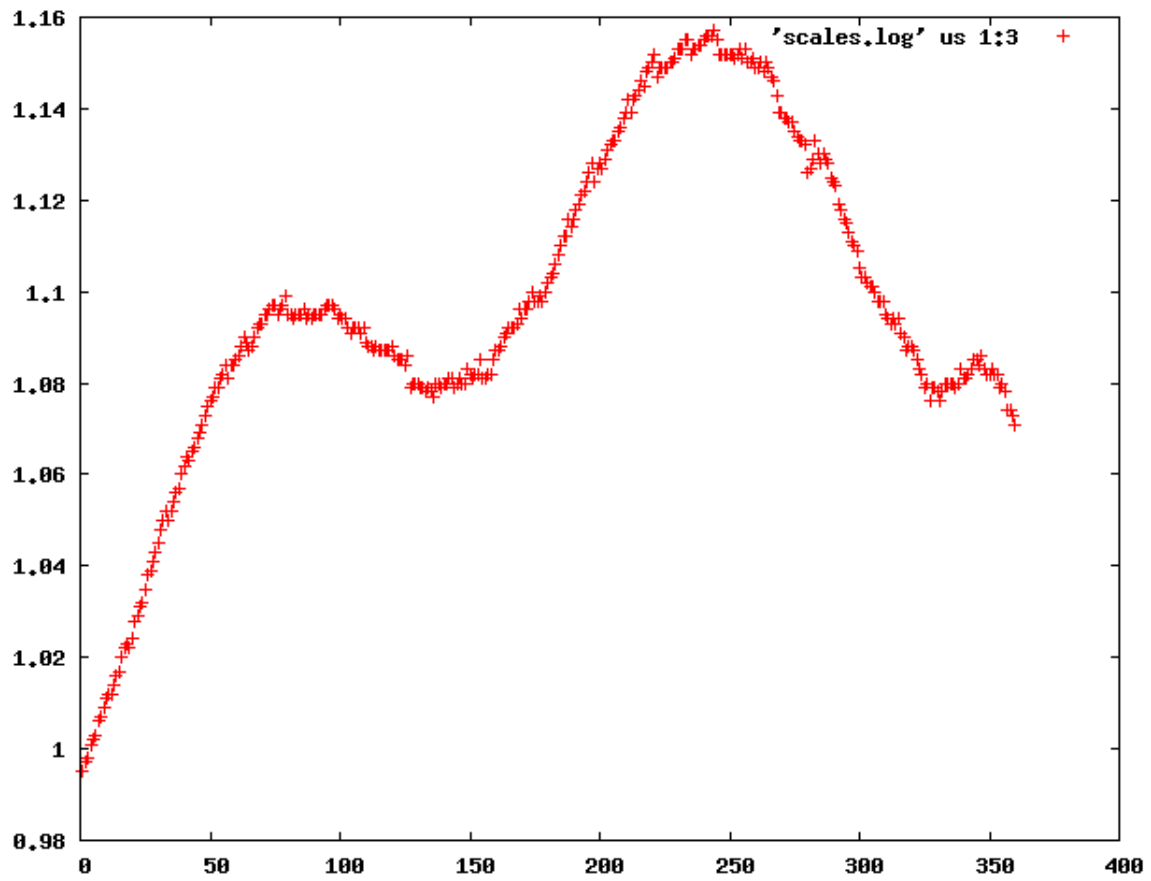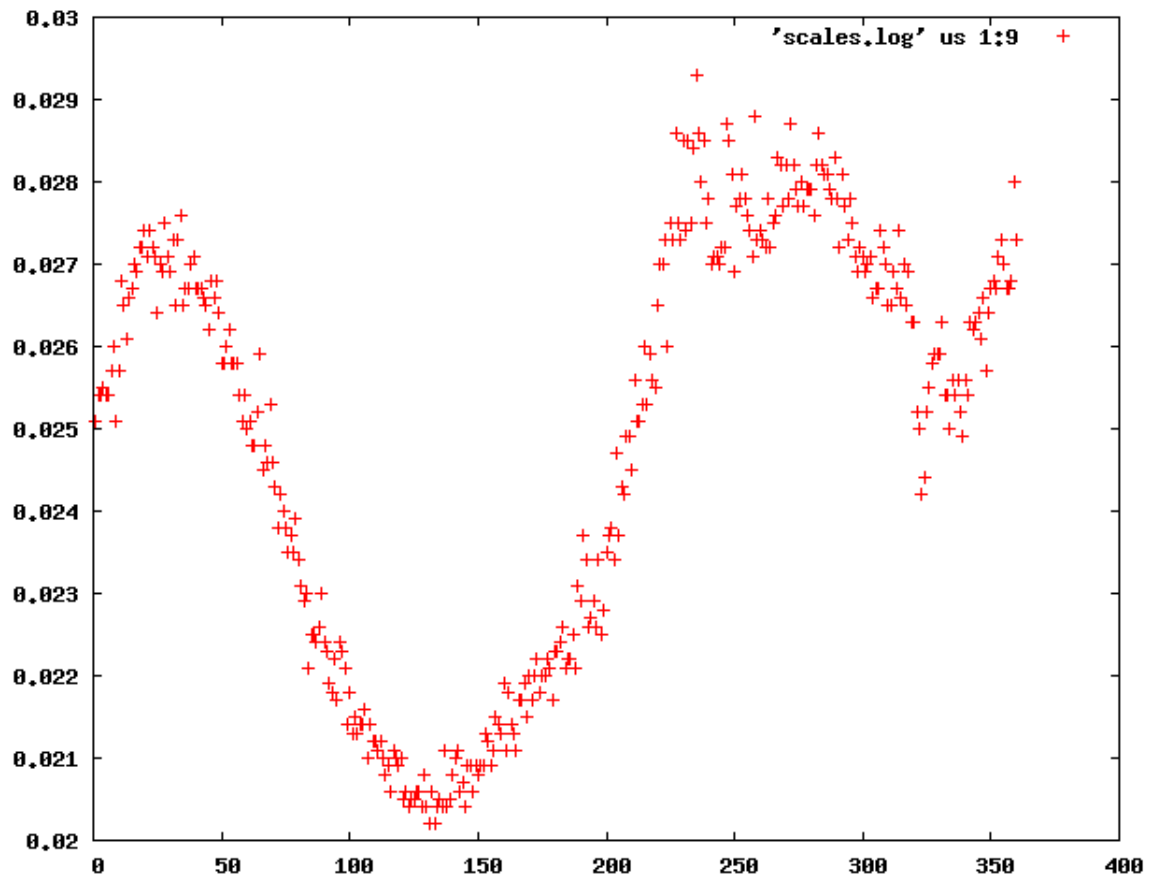
Next, we can use xdsstat to get frame-wise statistics from XDS_ASCII.HKL:
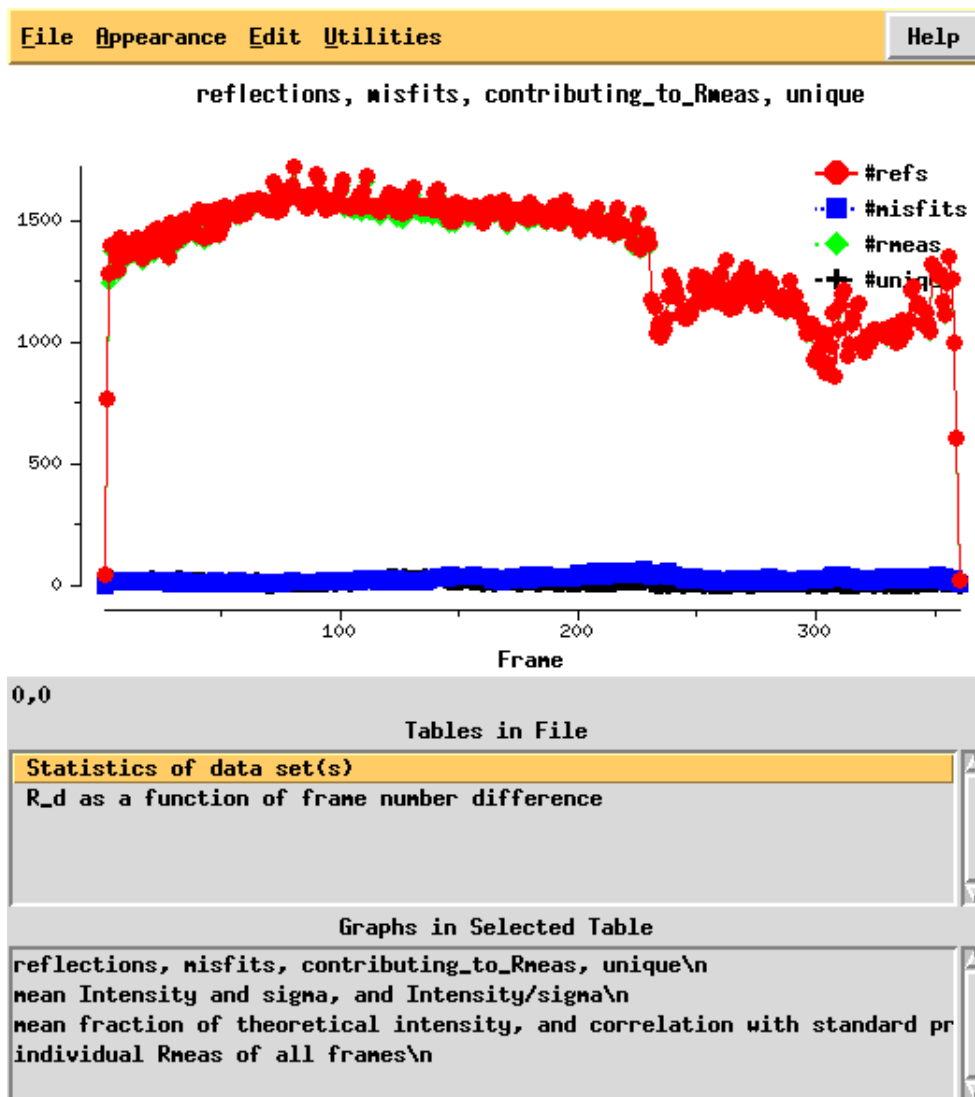
This shows a "jump" around frame 225 which is always bad for experimental phasing!

File Appearance Edit Utilities   Help

mean Intensity and sigma, and Intensity/sigma

344.4,301.4

Tables in File

Statistics of data set(s)
R_d as a function of frame number difference

Graphs in Selected Table

reflections, misfits, contributing_to_Rmeas, unique\n
mean Intensity and sigma, and Intensity/sigma\n
mean fraction of theoretical intensity, and correlation with standard pr
individual Rmeas of all frames\n

Around frame 225 the data are weakest, but they recover.

File Appearance Edit Utilities   Help

mean fraction of theoretical intensity, and correlation with standard profi
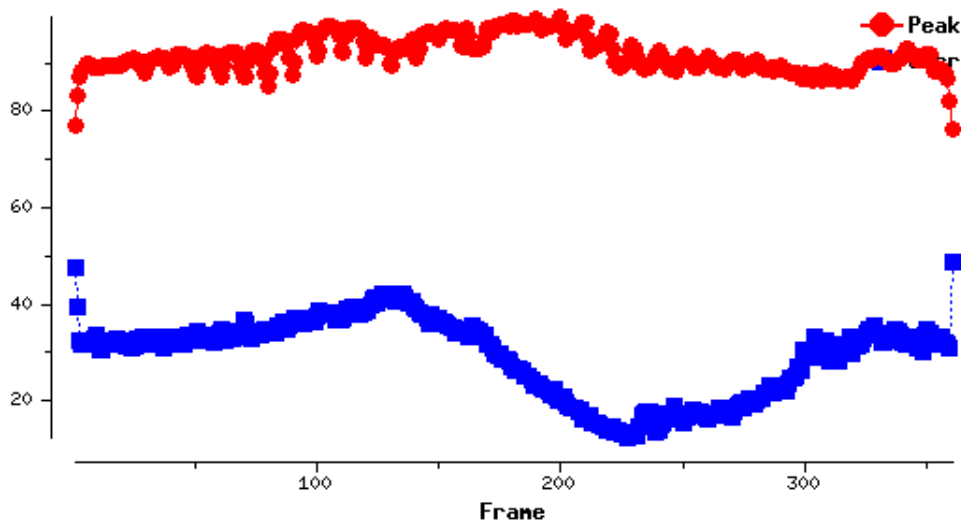
278.0,5.939

Tables in File

Statistics of data set(s)
R_d as a function of frame number difference

Graphs in Selected Table

reflections, misfits, contributing_to_Rmeas, unique\n
mean Intensity and sigma, and Intensity/sigma\n
mean fraction of theoretical intensity, and correlation with standard pr
individual Rmeas of all frames\n

In particular the correlation against standard profiles (blue curve) is *really* low.

File Appearance Edit Utilities   Help

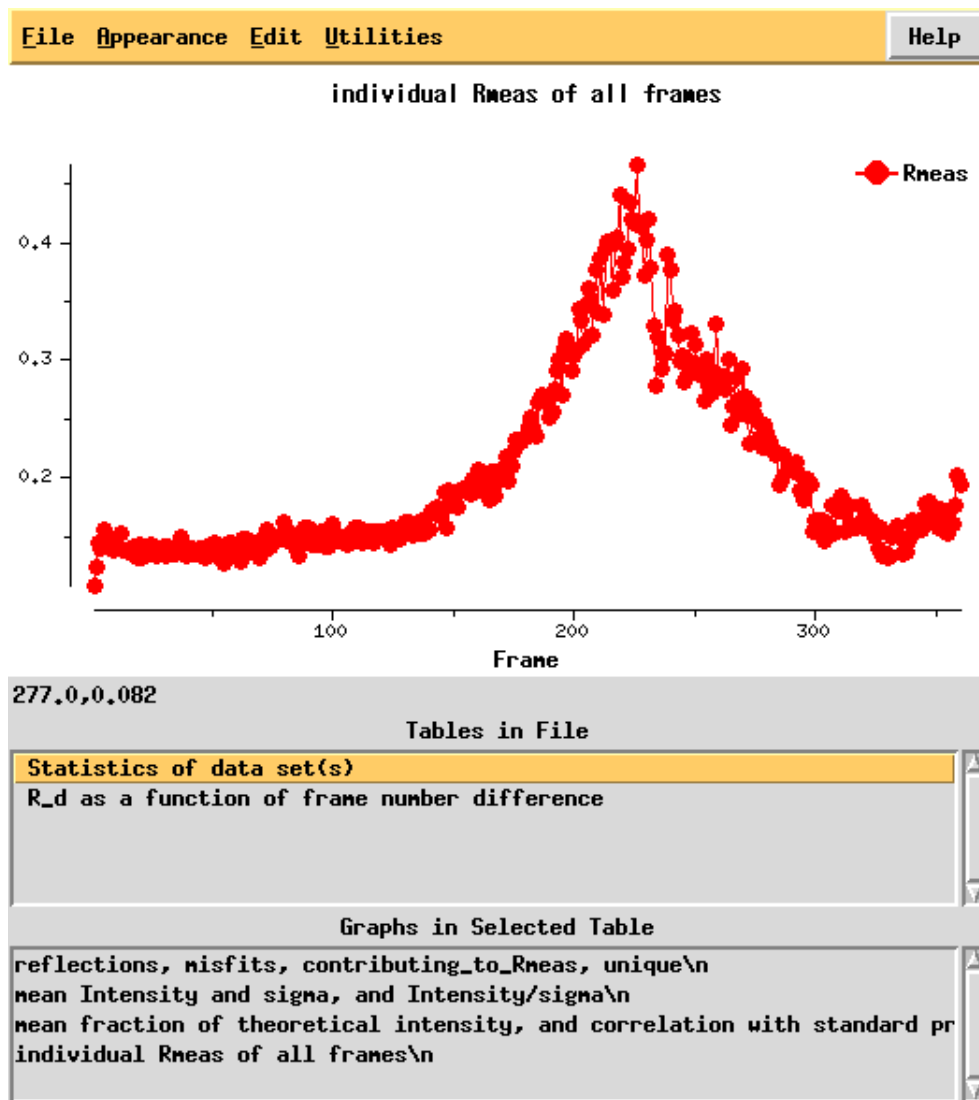### individual Rmeas of all frames

277.0,0.082

Tables in File

Statistics of data set(s)
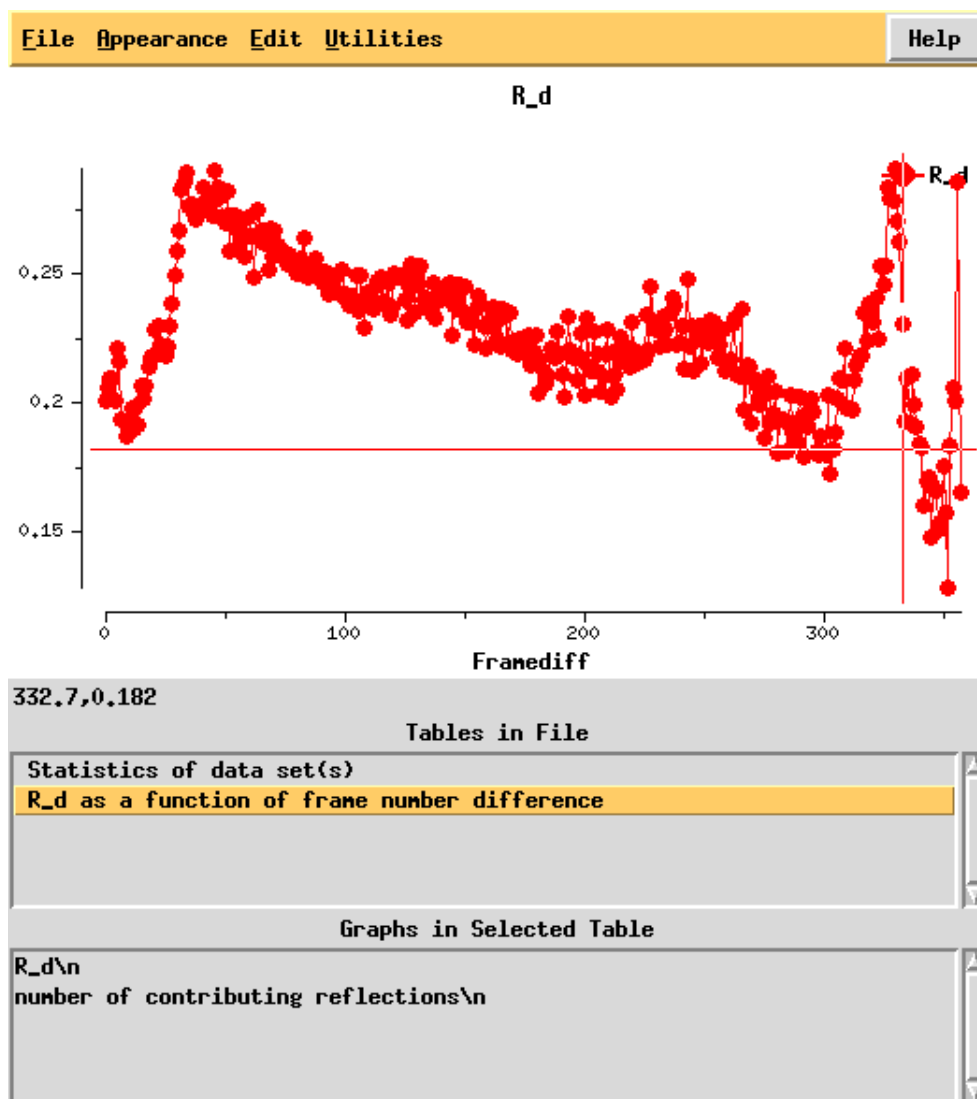R_d as a function of frame number difference

Graphs in Selected Table

reflections, misfits, contributing_to_Rmeas, unique\n
mean Intensity and sigma, and Intensity/sigma\n
mean fraction of theoretical intensity, and correlation with standard pr
individual Rmeas of all frames\n

R-factors peak around frame 225.

R_d helps to quantify radiation damage. Unfortunately, for this dataset this "R-factor as a function of frame number difference" behaves wildly, so we cannot use 0-dose extrapolation, like we successfully did for 1Y13.

High-remote

Due to a beamline problem, high-remote data collection stopped after 269 frames of 0.5° (the final frame is already affected). After restart of the beamline, another 100 frames were collected but they later turned out to merge badly with the first 269 frames - a hint that the monochromator was still heating up, or similar. So the latter frames were left out. The 269 frames are here (ftp://turn5.biologie.uni-konstanz.de/pub/datasets/3csl-hrem.tar) (1.4 Gb).

From CORRECT.LP :

```
┌──────────────────────────────────────────────────────────────────────────────┐
│    a          b          ISa                                                   │
│ 6.595E+00  3.032E-04    22.36                                                  │
│                                                                                │
│ ...                                                                            │
│                                                                                │
│      NOTE:       Friedel pairs are treated as different reflections.           │
│                                                                                │
│ SUBSET OF INTENSITY DATA WITH SIGNAL/NOISE >= -3.0 AS FUNCTION OF RESOLUTION    │
│ RESOLUTION      NUMBER OF REFLECTIONS    COMPLETENESS R-FACTOR  R-FACTOR COMPARED I/SIGMA   R-meas  Rmrgd │
│   LIMIT     OBSERVED  UNIQUE  POSSIBLE     OF DATA    observed  expected                          │
│                                                                                │
│    8.21      20187    6913     7245        95.4%       2.9%      3.3%    20065  27.89   3.6%    2. │
│    5.84      36272   12417    12782        97.1%       5.5%      5.3%    36116  17.46   6.8%    7. │
│    4.78      46716   16015    16500        97.1%       6.8%      6.5%    46473  14.43   8.4%   10. │
│    4.15      55299   18949    19484        97.3%       7.5%      7.2%    55003  13.00   9.3%   11. │
│    3.71      63751   21798    22065        98.8%      12.2%     12.2%    63371   8.51  15.1%   20. │
│    3.39      70787   24180    24422        99.0%      19.5%     20.0%    70378   5.58  24.0%   33. │
│    3.14      61197   25100    26452        94.9%      32.5%     34.2%    57925   2.88  41.3%   63. │
│    2.94      40481   21869    28566        76.6%      53.5%     57.8%    33568   1.42  72.1%  112. │
│    2.77      24584   16962    30228        56.1%      76.4%     82.4%    15055   0.77 107.5%  163. │
│   total     419274  164203   187744        87.5%      10.3%     10.4%   397954   8.06  12.9%   24. │
│                                                                                │
│                                                                                │
│ NUMBER OF REFLECTIONS IN SELECTED SUBSET OF IMAGES   428770                     │
│ NUMBER OF REJECTED MISFITS                             9102                     │
│ NUMBER OF SYSTEMATIC ABSENT REFLECTIONS                   0                     │
│ NUMBER OF ACCEPTED OBSERVATIONS                      419668                     │
│ NUMBER OF UNIQUE ACCEPTED REFLECTIONS                164343                     │
└──────────────────────────────────────────────────────────────────────────────┘
```

Inflection

360 frames (0.5° oscillation) at the inflection wavelength were collected after the peak data. They can be downloaded from here (ftp://turn5.biologie.uni-konstanz.de/pub/datasets/3csl-ip.tar) (1.8 Gb).

CORRECT.LP has:

```
┌──────────────────────────────────────────────────────────────────────────────┐
│    a          b          ISa                                                   │
│ 6.514E+00  5.329E-04    16.97                                                  │
│                                                                                │
│ ...                                                                            │
│                                                                                │
│      NOTE:       Friedel pairs are treated as different reflections.           │
│                                                                                │
│ SUBSET OF INTENSITY DATA WITH SIGNAL/NOISE >= -3.0 AS FUNCTION OF RESOLUTION    │
│ RESOLUTION      NUMBER OF REFLECTIONS    COMPLETENESS R-FACTOR  R-FACTOR COMPARED I/SIGMA   R-meas  Rmrgd │
│   LIMIT     OBSERVED  UNIQUE  POSSIBLE     OF DATA    observed  expected                          │
│                                                                                │
│    8.28      26530    7039     7111        99.0%       4.4%      4.6%    26516  23.79   5.2%    3. │
│    5.90      48700   12589    12598        99.9%       8.7%      8.5%    48700  14.00  10.2%    9. │
│    4.83      62546   16177    16208        99.8%      12.0%     11.5%    62546  10.85  13.9%   14. │
│    4.18      72644   19076    19117        99.8%      13.6%     13.1%    72637   9.64  15.9%   17. │
│    3.75      80829   21710    21740        99.9%      23.8%     24.0%    80829   5.92  27.9%   32. │
│    3.42      86652   23874    23917        99.8%      38.5%     39.9%    86652   3.71  45.2%   53. │
│    3.17      73630   25264    26115        96.7%      64.4%     68.0%    71837   1.82  78.5%  109. │
│    2.96      48079   22582    28004        80.6%      99.2%    107.3%    43325   0.86 129.6%  186. │
│    2.80      28417   17801    29828        59.7%     155.6%    168.5%    20099   0.43 214.2%  307. │
│   total     528027  166112   184638        90.0%      17.6%     17.8%   513141   5.98  20.9%   38. │
│                                                                                │
│                                                                                │
│ NUMBER OF REFLECTIONS IN SELECTED SUBSET OF IMAGES   534898                     │
│ NUMBER OF REJECTED MISFITS                             6486                     │
│ NUMBER OF SYSTEMATIC ABSENT REFLECTIONS                   0                     │
│ NUMBER OF ACCEPTED OBSERVATIONS                      528412                     │
│ NUMBER OF UNIQUE ACCEPTED REFLECTIONS                166224                     │
└──────────────────────────────────────────────────────────────────────────────┘
```
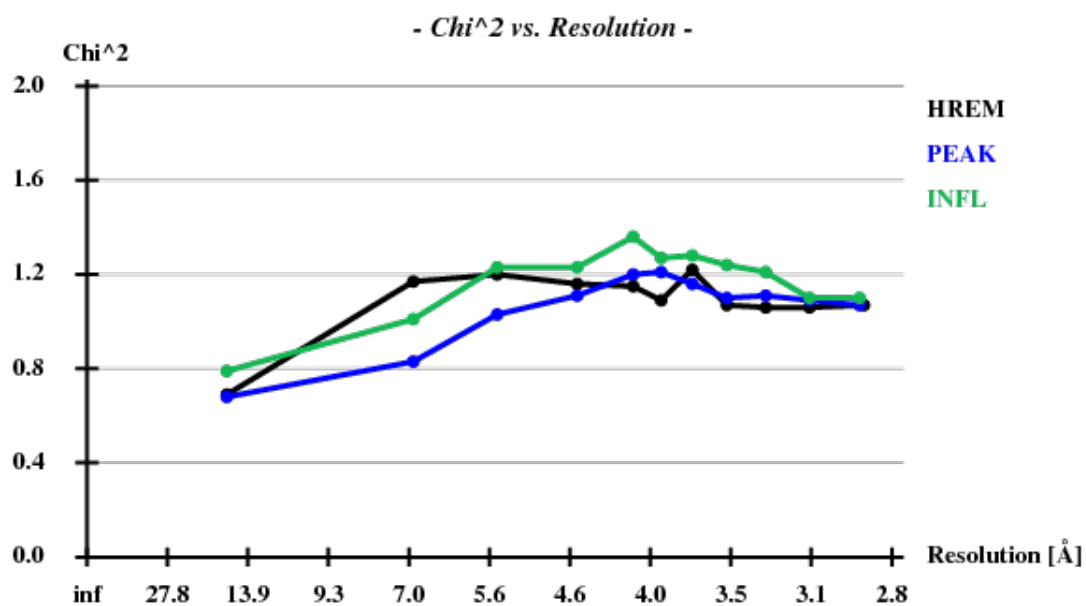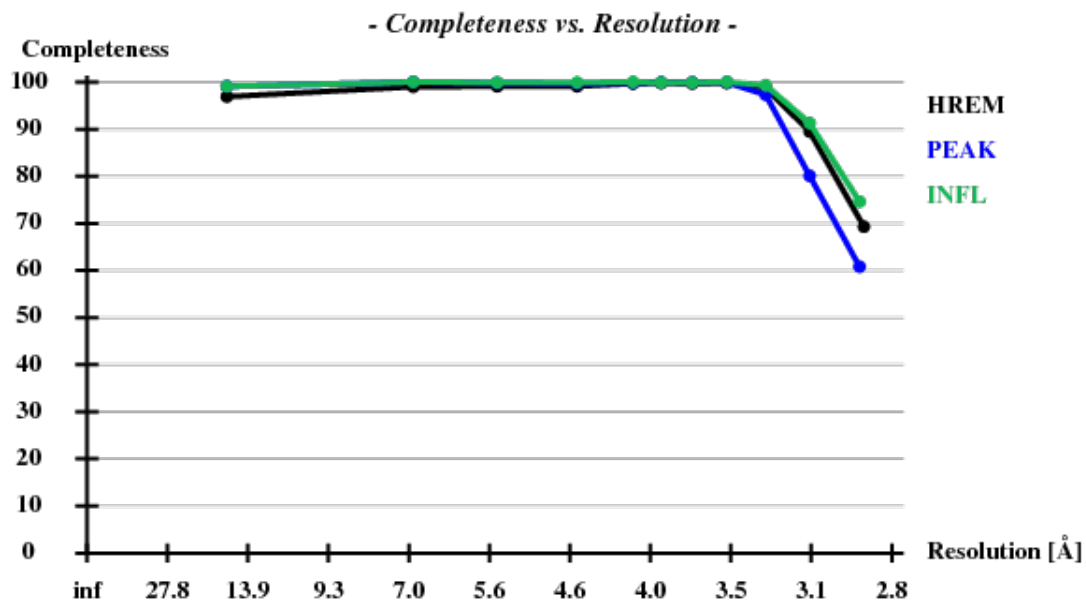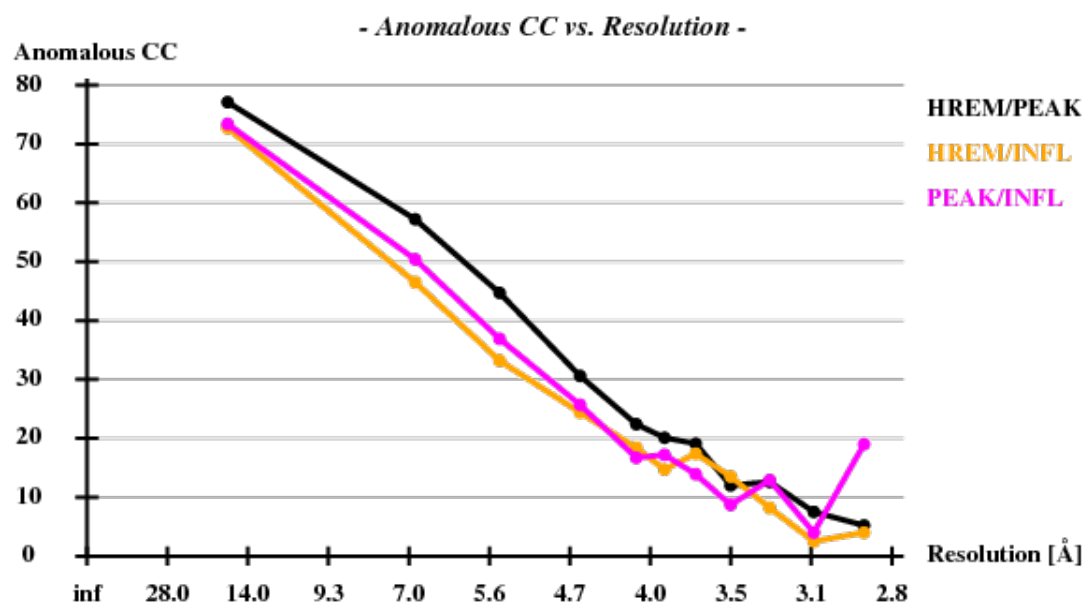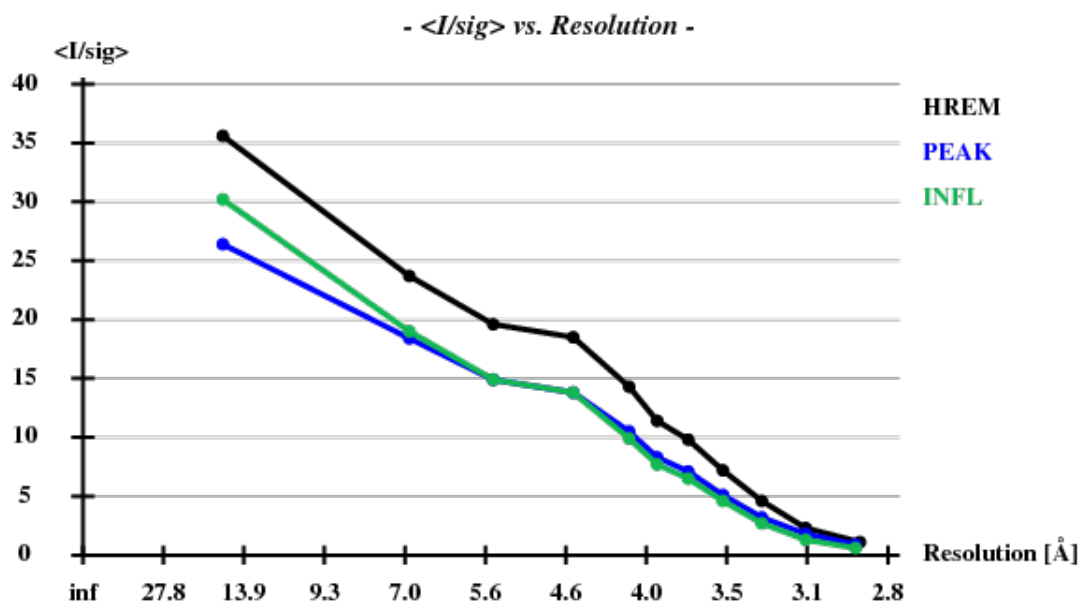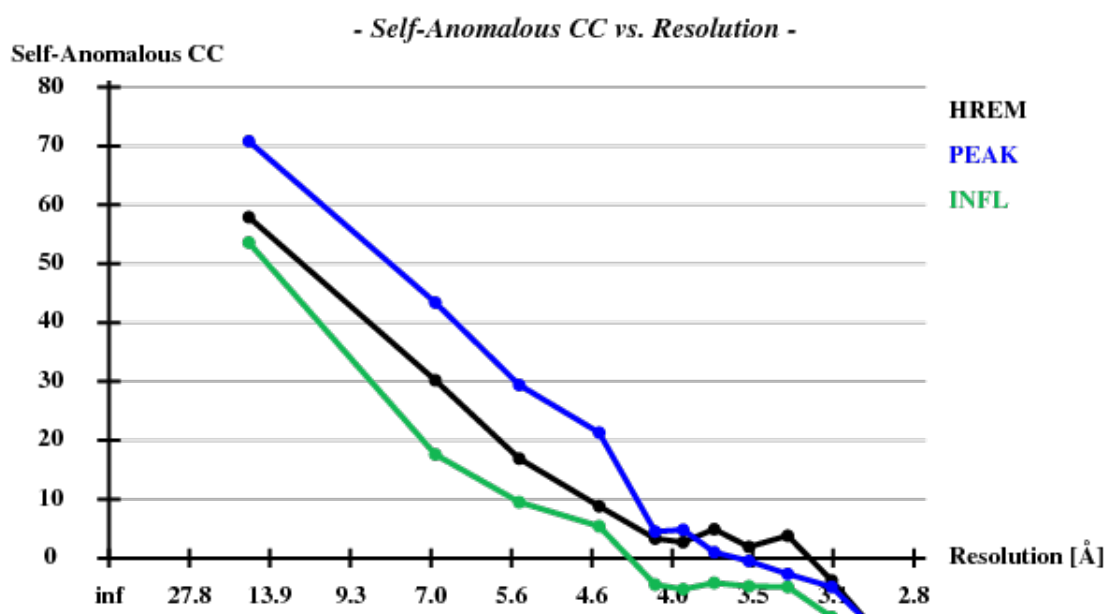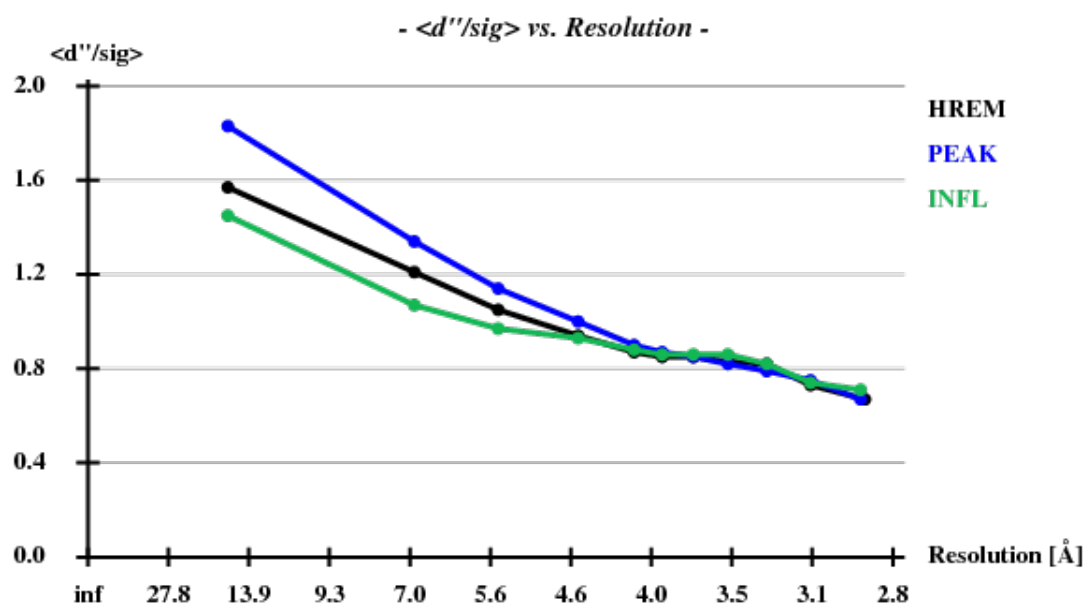
Scaling

Structure solution
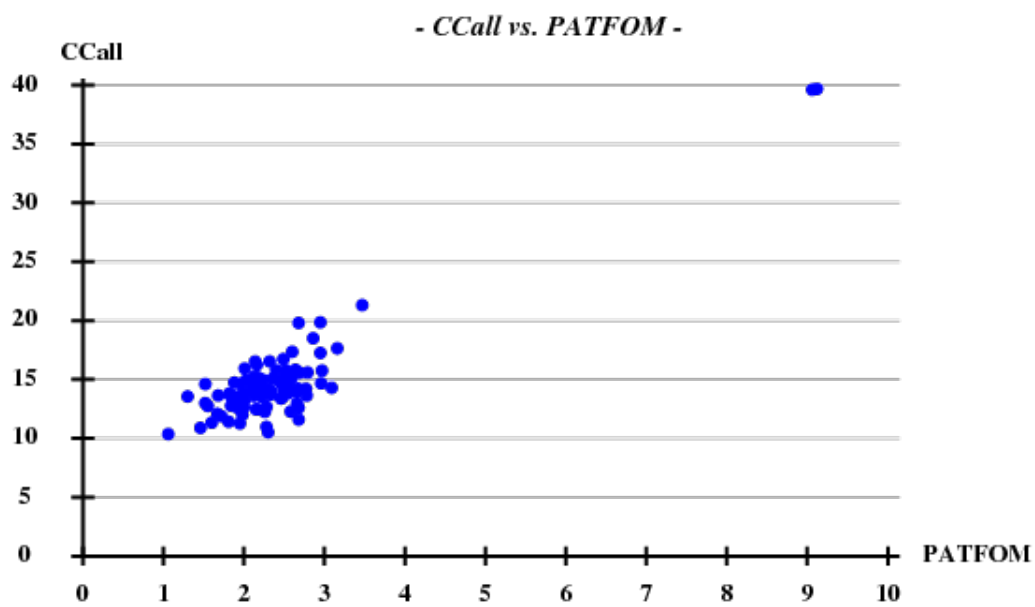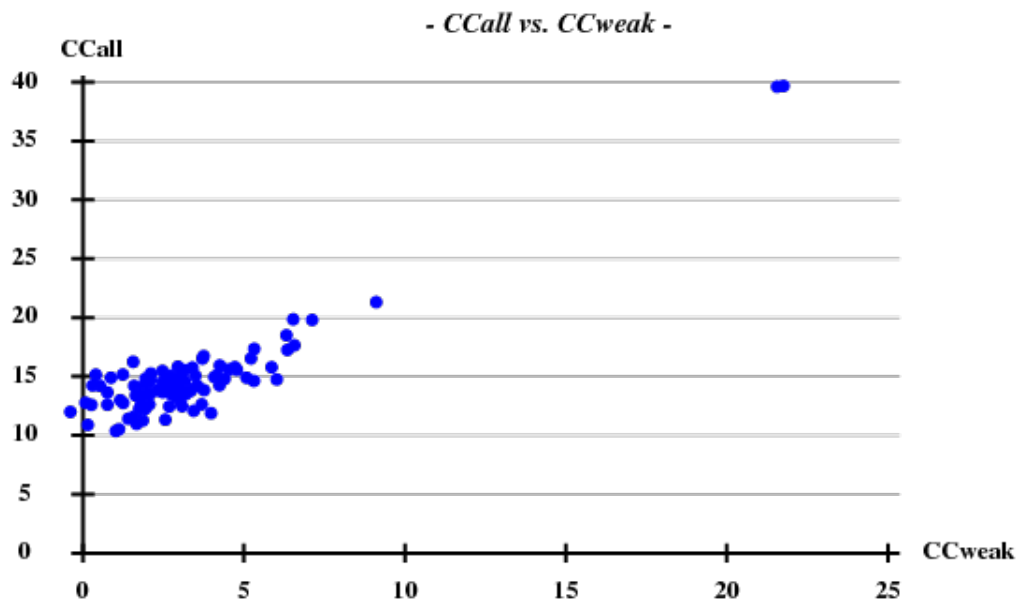
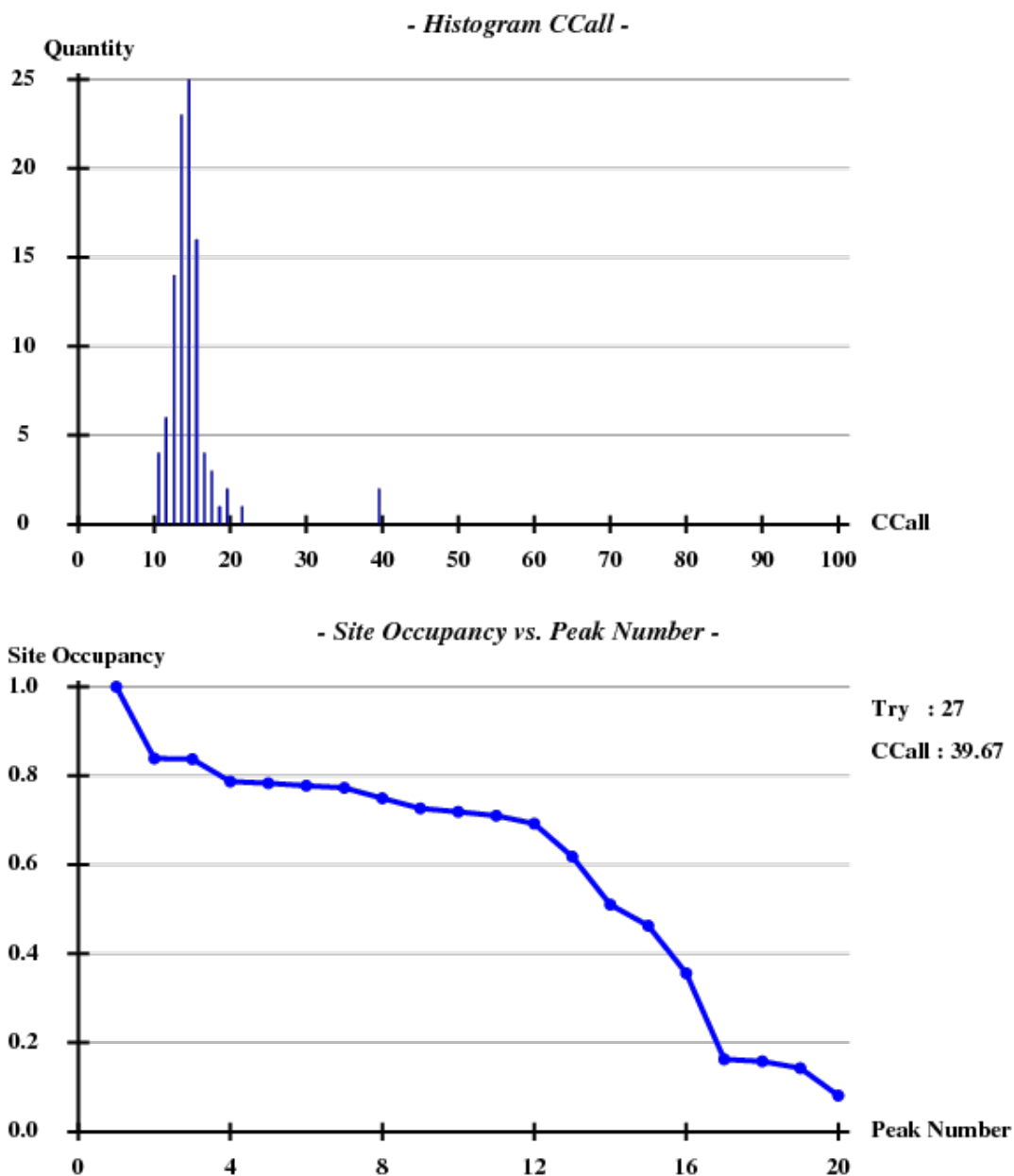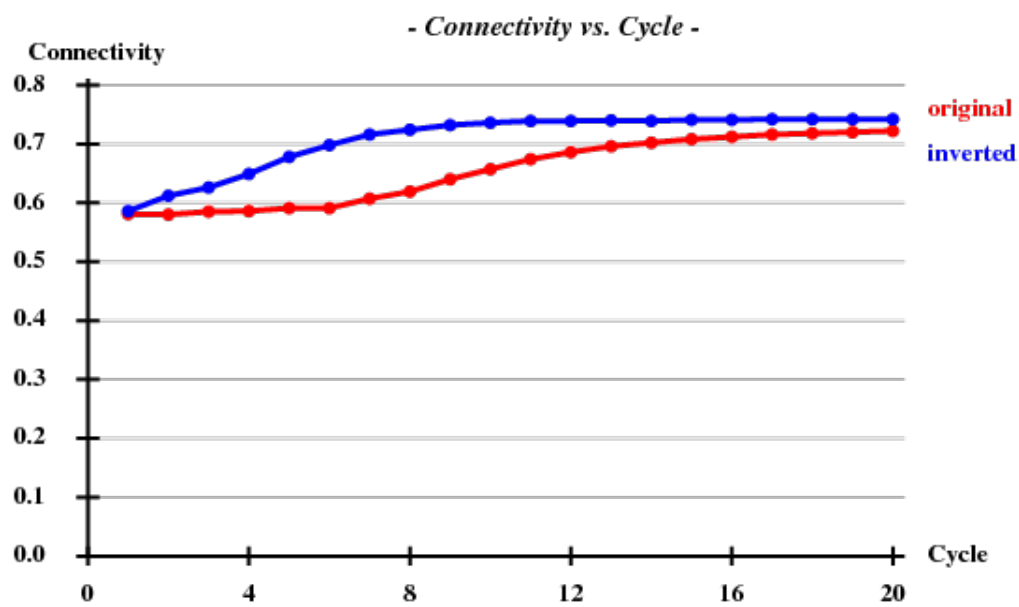We use hkl2map for solving the structure.

SHELXC statistics

- Completeness vs. Resolution -

- Chi^2 vs. Resolution -

- *<I/sig> vs. Resolution -*

<I/sig>

HREM
PEAK
INFL

Resolution [Å]

inf   27.8   13.9   9.3   7.0   5.6   4.6   4.0   3.5   3.1   2.8

- *Anomalous CC vs. Resolution -*

Anomalous CC

HREM/PEAK
HREM/INFL
PEAK/INFL

Resolution [Å]

inf   28.0   14.0   9.3   7.0   5.6   4.7   4.0   3.5   3.1   2.8

- *<d"/sig> vs. Resolution -*

- *Self-Anomalous CC vs. Resolution -*

SHELXD statistics

- CCall vs. CCweak -

- CCall vs. PATFOM -

- Histogram CCall -

Quantity

- Site Occupancy vs. Peak Number -
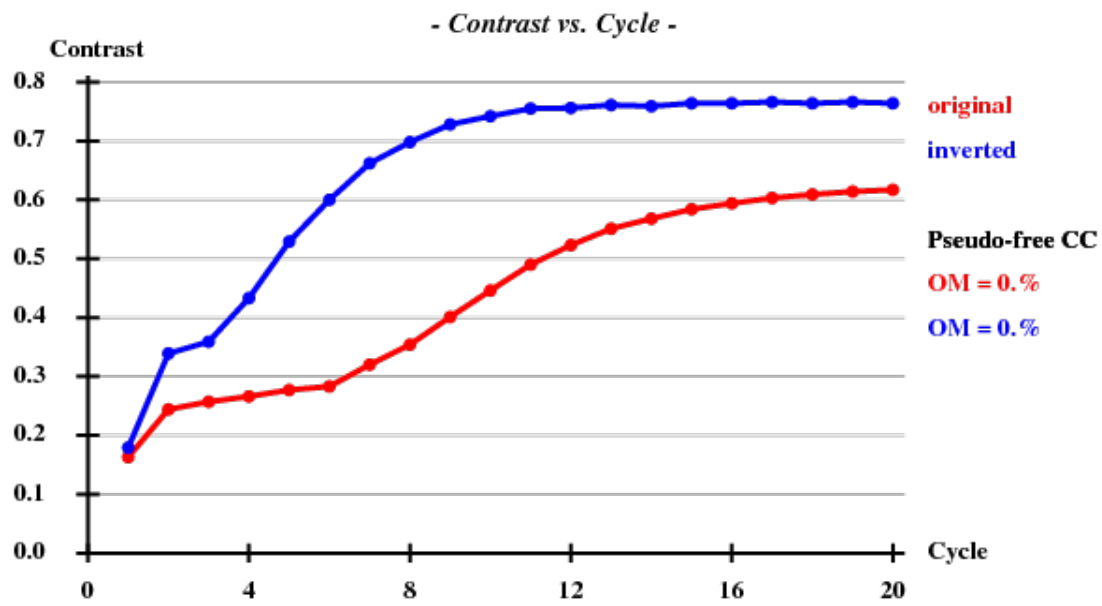
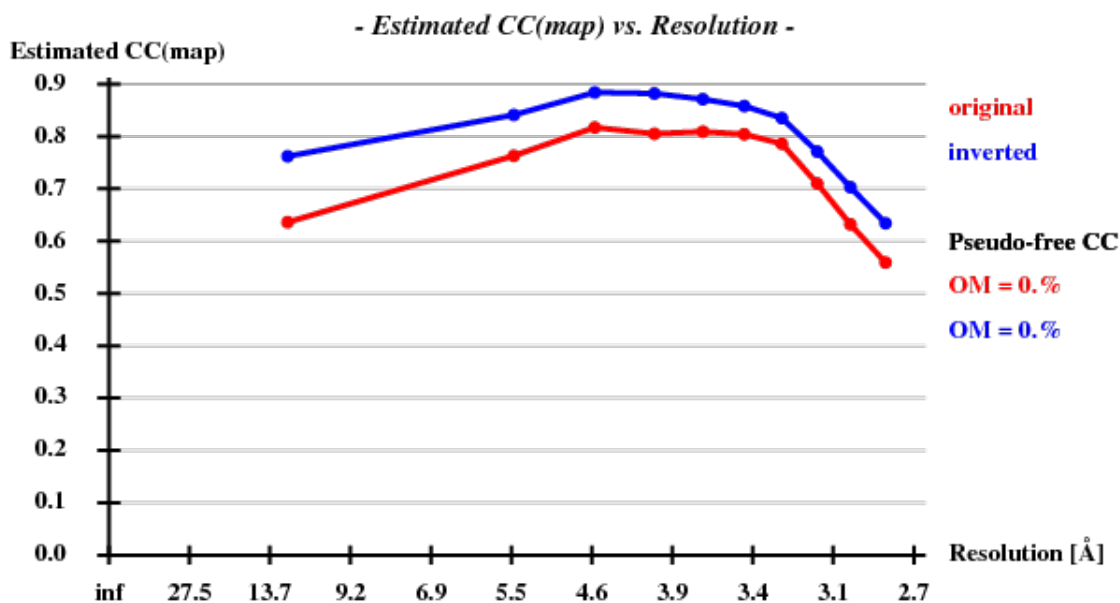**Site Occupancy**

Try : 27

CCall : 39.67

SHELXE statistics

Since there are 1852 residues in the ASU, the solvent content is about 72.5%.
The correct hand ("inverted") becomes immediately clear - it is superior in all
respects than the "original" hand.

- Contrast vs. Cycle -

Contrast

original
inverted

Pseudo-free CC
OM = 0.%
OM = 0.%

Cycle

- Connectivity vs. Cycle -

Connectivity

original
inverted

Cycle

**- Estimated CC(map) vs. Resolution -**

Then we try to get a poly-ala backbone tracing, using the 16 Se sites found by SHELXD:

```
shelxe.beta -a -q -m15 -s0.725 -b -h16 -n2 -i mad mad_fa
```

This does not yield a complete chain, but rather about 50% of it, and the CC is slightly less than 20%, so that we cannot consider the structure as solved yet. However, the phases are good enough for finding 2 additional sites. We iterate this, and finally the Se sites all have a density of at least 20 sigma. The model from shelxe as well as the deposited structure are shown below:

Successfully read coordinates file coot-download/3csl.pdb. Molecule number 1 created.
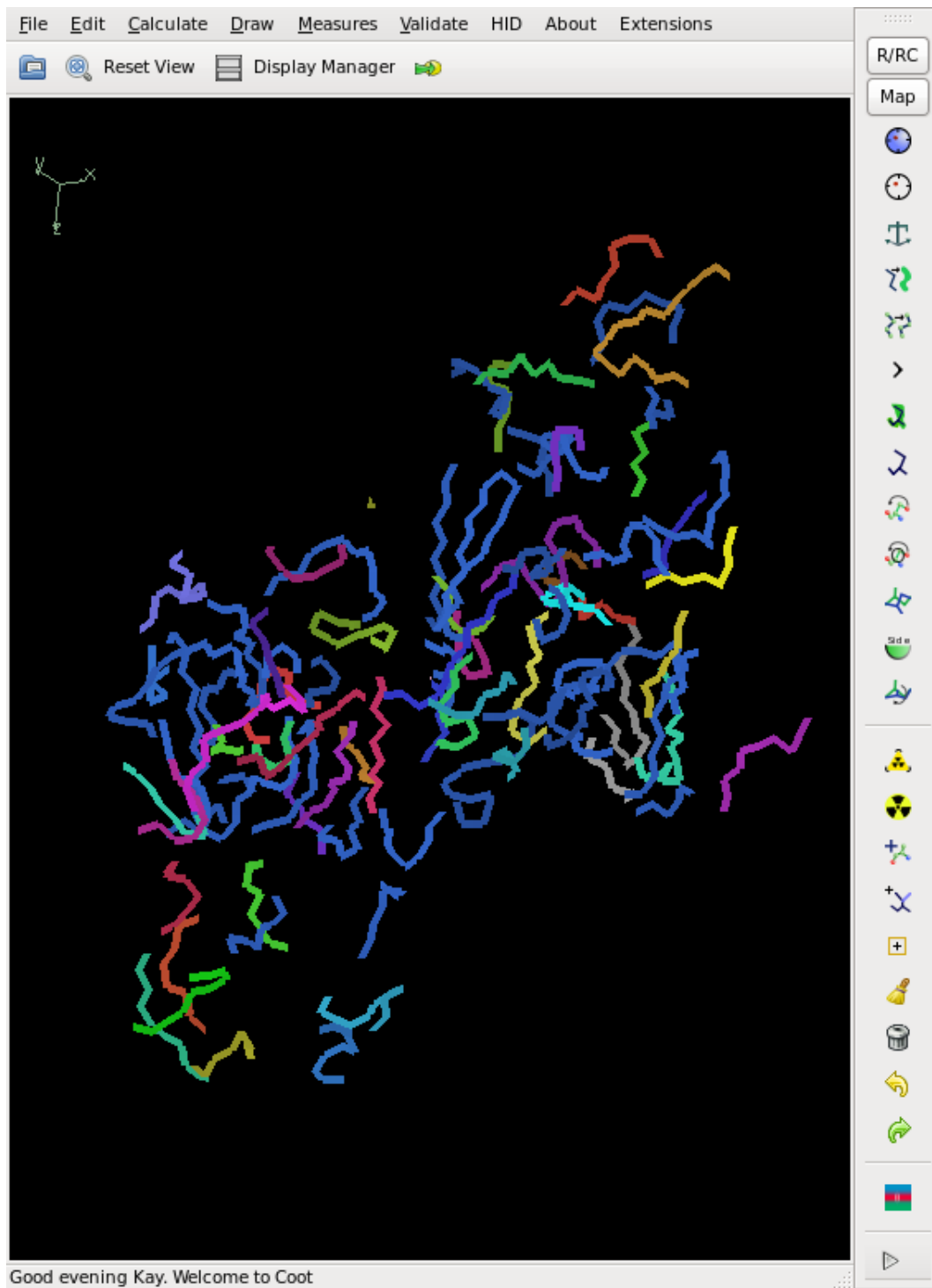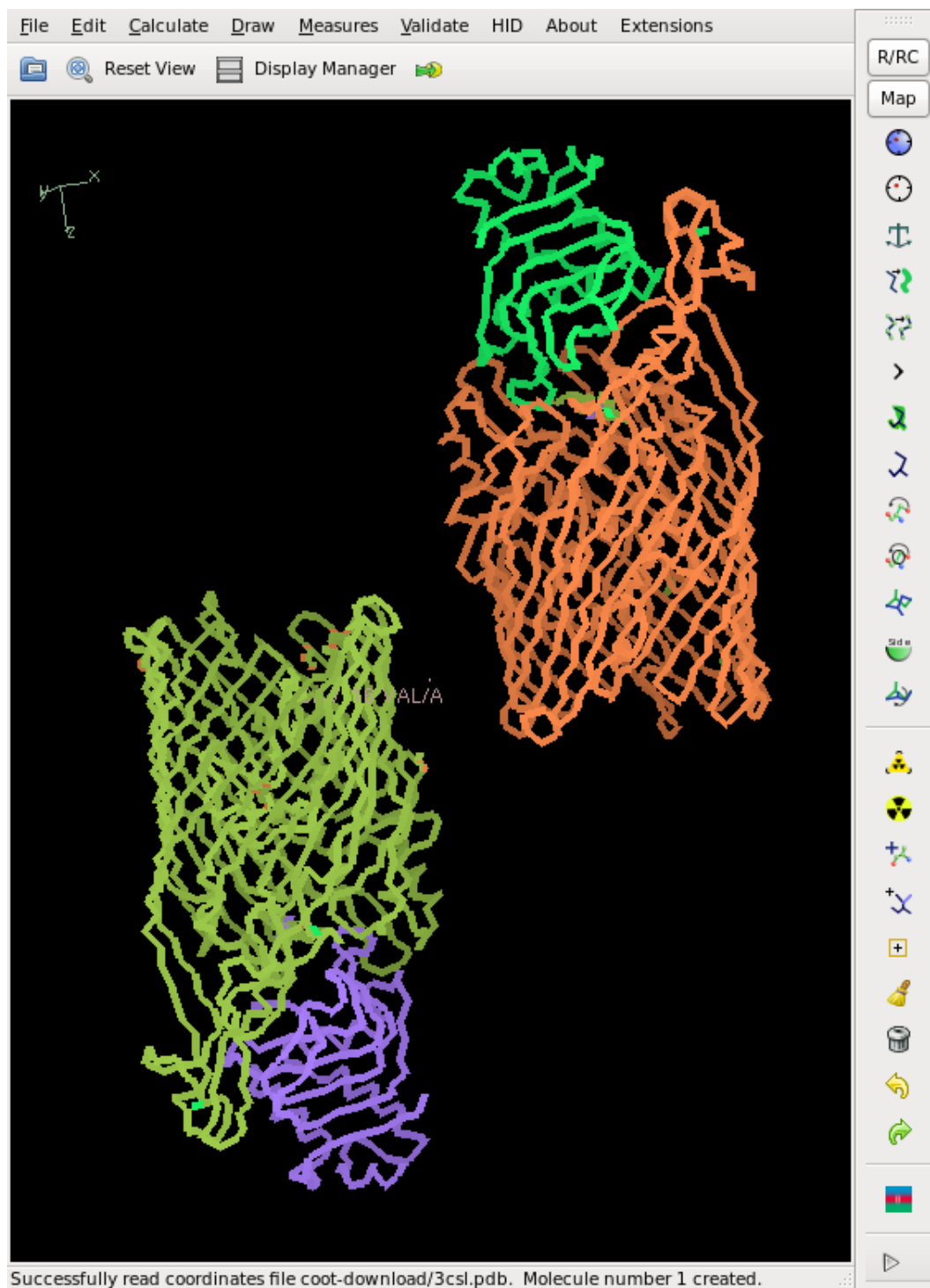
No doubt that one can solve the structure from here, maybe after HA refinement with SHARP, and model building with buccanneer or Arp/wArp.

Availability of data

There are files with amplitudes (3csl-pk-F.mtz, 3csl-rh-F.mtz, 3csl-ip-F.mtz) and intensities (3csl-pk-I.mtz, 3csl-rh-I.mtz, 3csl-ip-I.mtz) as well as mad_i.pdb and mad_i.phs (written by SHELXE) available from [1] (ftp://turn5.biologie.uni-konstanz.de/pub/xds-datared/3csl/) .

additional information for those who want to complete the structure

These are the entire sequences of HasR and HasA - before solving the structure it was not known that the N-terminus of HasR was disordered.

```
AQAEASSAQAAQQKNFNIAAQPLQSAMLRFAEQAGMQVFFDEVKLDGMQAAALNGSMSVEQGLRRLIGGNPVAFRLQPQGQIVLSRLPTANGDGGALALD
SLTVLGAGGNNANDWVYDEPRSVSVISREQMDNRPARHAADILEQTTGAYSSVSQQDPALSVNIRGIQDYGRVNMNIDGMRQNFQKSGHGQRNGTMYIDS
ELLSGVTIDKGTTGGMGSAGTLGGIATFNTVSASDFLAPGKELGGKLHASTGDNGTHFIGSGILALGNETGDILLAASERHLGDYWPGNKGDIGNIRINN
DTGNYDRYAESIKNNKIPDTHYRMHSRLAKVGWNLPANQRLQLSYLQTQTASPIAGTLTNLGTRPPYELGWKRTGYTDVMARNAAFDYSLAPEDVDWLDF
QAKLYYVDTQDDSDTYSTSSLLDNGYATRTRLRTYGAQAQNTSRFSLAPGHDFRANYGLEFYYDKATSDSSRQGMEGVTPAGNRSVASLFANLTYDYDGW
LTLEGGLRYDRYRLRGQTGLSYPDLAKDGQRYTIDNPCKALRLTGCSTTTREDWDVDRDQGKLSPTLAVAVRPGVEWLELYTTYGKSWRPPAITETLTNG
SAHSSSTQYPNPFLQPERSRAWEVGFNVQQPDLWFEGDRLVAKVaYFDTKVDNYINLAIDRNKPGLVQPSIGNAAYVNNLSKTRFRGLEYQLNYDAGVFY
ADLTYTHMIGKNEFCSNKAWLGGRLRYGDGSRRGNFYVEPDAASNDFVTCDGGTQFGSAAYLPGDRGSVTLGGRAFDRKLDAGVTVRFAPGYQDSSVPSN
YPYLADWPKYTLFDLYASYKLTDSLTLRGSVENLTNRAYVVSYGETLANTLGRGRTVQGGVEYRF

MRGSHHHHHHGIRMRARYPAFSVNYDSSFGGYSIHDYLGQWASTFGDVNH
TNGNVTDANSGGFYGGSLSGSQYAISSTANQVTAFVAGGNLTYTLFNEPA
HTLYGQLDSLSFGDGLSGGDTSPYSIQVPDVSFGGLNLSSLQAQGHDGVV
HQVVYGLMSGDTGALETALNGILDDYGLSVNSTFDQVAAATAVGVQHADS
PELLAA
```

Retrieved from "http://strucbio.biologie.uni-konstanz.de/xdswiki/index.php/3CSL"

- This page was last modified on 18 March 2011, at 21:08.