2QVO.xds

From XDSwiki

(Redirected from 2QVO)

This is an example of S-SAD structure solution (PDB id 2QVO (http://www.rcsb.org /pdb/explore.do?structureId=2QVO)), a 95-residue protein used by James Tucker Swindell II to establish optimized procedures for data reduction. The data available to solve the structure are two runs of 360° collected at a wavelength of 1.9Å.

Contents

XDS data reduction

In the course of writing this up, it turned out that it was not necessary to scale the two datasets together, using XSCALE, because the structure can be solved from any of the two, separately. But, of course, structure solution would be easier when merging the data (try for yourself!).

dataset 1

Using "generate_XDS.INP ../../APS/22-ID/2qvo/ACA10_AF1382_1.0???" we obtain:

```
JOB= XYCORR INIT COLSPOT IDXREF DEFPIX INTEGRATE CORRECT
ORGX= 1996.00 ORGY= 2028.00  ! check these values with adxv !
DETECTOR_DISTANCE= 125.000
OSCILLATION_RANGE= 1.000
X-RAY_WAVELENGTH= 1.90000
NAME_TEMPLATE_OF_DATA_FRAMES=../../APS/22-ID/2qvo/ACA10_AF1382_1.0???
! REFERENCE_DATA_SET=xxx/XDS_ASCII.HKL ! e.g. to ensure consistent indexing
DATA_RANGE=1 360
SPOT_RANGE=1 180
! BACKGROUND_RANGE=1 10 ! rather use defaults (first 5 degree of rotation)

SPACE_GROUP_NUMBER=0                     ! 0 if unknown
UNIT_CELL_CONSTANTS= 70 80 90 90 90 90 ! put correct values if known
INCLUDE_RESOLUTION_RANGE=50 0  ! after CORRECT, insert high resol limit; re-run CORRECT

FRIEDEL'S_LAW=FALSE     ! This acts only on the CORRECT step
! If the anom signal turns out to be, or is known to be, very low or absent,
! use FRIEDEL'S_LAW=TRUE instead (or comment out the line); re-run CORRECT

! remove the "!" in the following line:
! STRICT_ABSORPTION_CORRECTION=TRUE
! if the anomalous signal is strong: in that case, in CORRECT.LP the three
! "CHI^2-VALUE OF FIT OF CORRECTION FACTORS" values are significantly> 1, e.g. 1.5
!
! exclude (mask) untrusted areas of detector, e.g. beamstop shadow :
! UNTRUSTED_RECTANGLE= 1800 1950 2100 2150 ! x-min x-max y-min y-max ! repeat
! UNTRUSTED_ELLIPSE= 2034 2070 1850 2240 ! x-min x-max y-min y-max ! if needed
!
! parameters with changes wrt default values:
TRUSTED_REGION=0.00 1.2  ! partially use corners of detectors; 1.41421=full use
VALUE_RANGE_FOR_TRUSTED_DETECTOR_PIXELS=7000. 30000. ! often 8000 is ok
MINIMUM_ZETA=0.05        ! integrate close to the Lorentz zone; 0.15 is default
STRONG_PIXEL=6           ! COLSPOT: only use strong reflections (default is 3)
MINIMUM_NUMBER_OF_PIXELS_IN_A_SPOT=3 ! default of 6 is sometimes too high
REFINE(INTEGRATE)=CELL BEAM ORIENTATION ! AXIS DISTANCE

! parameters specifically for this detector and beamline:
DETECTOR= CCDCHESS MINIMUM_VALID_PIXEL_VALUE= 1 OVERLOAD= 65500
NX= 4096 NY= 4096  QX= .0732420000  QY= .0732420000 ! to make CORRECT happy if frames are unavailable
DIRECTION_OF_DETECTOR_X-AXIS=1 0 0
DIRECTION_OF_DETECTOR_Y-AXIS=0 1 0
INCIDENT_BEAM_DIRECTION=0 0 1
ROTATION_AXIS=1 0 0    ! at e.g. SERCAT ID-22 this needs to be -1 0 0
FRACTION_OF_POLARIZATION=0.98   ! better value is provided by beamline staff!
POLARIZATION_PLANE_NORMAL=0 1 0
```

Now we run "xds_par". This runs to completion. We should at least inspect, using XDS-Viewer, the file FRAME.cbf since this shows us the last frame of the dataset, with boxes superimposed which correspond to the expected locations of reflections.

The automatic spacegroup determination (CORRECT.LP) comes up with

```
 LATTICE-  BRAVAIS-   QUALITY  UNIT CELL CONSTANTS (ANGSTROEM & DEGREES)    REINDEXING TRANSFORMATION
CHARACTER  LATTICE     OF FIT      a      b      c   alpha  beta gamma

*  44        aP          0.0      41.2   53.5   53.5  90.3  90.1  90.1   -1  0  0  0  0  1  0  0  0  0 -1  0
*  31        aP          0.8      41.2   53.5   53.5  89.7  90.1  89.9    1  0  0  0  0  1  0  0  0  0  1  0
*  25        mC          1.4      75.4   75.8   41.2  90.0  90.1  90.0    0  1 -1  0  0 -1 -1  0 -1  0  0  0
*  35        mP          1.8      53.5   41.2   53.5  90.1  90.3  90.1    0 -1  0  0  1  0  0  0  0  0  1  0
*  23        oC          3.1      75.4   75.8   41.2  90.0  90.1  90.0    0  1 -1  0  0 -1 -1  0 -1  0  0  0
*  20        mC          3.9      75.8   75.4   41.2  90.1  90.0  90.0    0  1  1  0  0  1 -1  0 -1  0  0  0
*  34        mP          5.1      41.2   53.5   53.5  90.3  90.1  90.1    1  0  0  0  0  0  1  0  0 -1  0  0
*  33        mP          5.3      41.2   53.5   53.5  90.3  90.1  90.1   -1  0  0  0  0  1  0  0  0  0 -1  0
*  32        oP          6.1      41.2   53.5   53.5  90.3  90.1  90.1   -1  0  0  0  0  1  0  0  0  0 -1  0
*  21        tP          7.3      53.5   53.5   41.2  90.1  90.1  90.3    0  1  0  0  0  0 -1  0 -1  0  0  0
   39        mC        249.8     114.5   41.2   53.5  90.1  90.3  69.0    1 -2  0  0  1  0  0  0  0  0  1  0
```

indicating at most tetragonal symmetry. Below this table, CORRECT calculates R-factors for each of the lattices whose metric symmetry is compatible with the cell of the crystal (marked by * in the table above):

SPACE-GROUP NUMBER	UNIT CELL CONSTANTS a	b	c	alpha	beta	gamma	UNIQUE	Rmeas	COMPARED	LATTICE-CHARACTER
5	75.8	75.4	41.2	90.0	90.0	90.0	900	40.8	5882	20 mC
* 75	53.5	53.5	41.2	90.0	90.0	90.0	469	8.4	6313	21 tP
89	53.5	53.5	41.2	90.0	90.0	90.0	282	39.2	6500	21 tP
21	75.4	75.8	41.2	90.0	90.0	90.0	506	39.8	6276	23 oC
5	75.4	75.8	41.2	90.0	90.1	90.0	901	40.7	5881	25 mC
1	41.2	53.5	53.5	89.7	90.1	89.9	1699	8.2	5083	31 aP
16	41.2	53.5	53.5	90.0	90.0	90.0	521	39.8	6261	32 oP
3	53.5	41.2	53.5	90.0	90.3	90.0	931	8.2	5851	35 mP
3	41.2	53.5	53.5	90.0	90.1	90.0	918	40.7	5864	33 mP
3	41.2	53.5	53.5	90.0	90.1	90.0	918	40.9	5864	34 mP
1	41.2	53.5	53.5	90.3	90.1	90.1	1699	8.2	5083	44 aP

thus suggesting spacegroup #75 but we should know that this does not take screw axes into account. Therefore we use "pointless xdsin XDS_ASCII.HKL" and are told that this is actually spacegroup P4_2 (# 77). Alternatively, we could have inspected the list further down in CORRECT.LP:

REFLECTIONS OF TYPE H,0,0 0,K,0 0,0,L OR EXPECTED TO BE ABSENT (*)

H	K	L	RESOLUTION	INTENSITY	SIGMA	INTENSITY/SIGMA	#OBSERVED
0	0	1	41.248	0.8487E+01	0.1339E+01	6.34	4
0	0	3	13.749	-0.7977E-03	0.3786E+01	0.00	4
0	0	4	10.312	0.1305E+06	0.4660E+04	27.99	1
0	0	5	8.250	0.1318E+01	0.6316E+01	0.21	4
0	0	6	6.875	0.2939E+05	0.5284E+03	55.61	4
0	0	7	5.893	0.5439E+01	0.9235E+01	0.59	4
0	0	8	5.156	0.1298E+05	0.2371E+03	54.73	4
0	0	9	4.583	0.3308E+02	0.1327E+02	2.49	4
0	0	10	4.125	0.3809E+05	0.6867E+03	55.47	4
0	0	11	3.750	-0.1987E+02	0.1767E+02	-1.12	4
0	0	12	3.437	0.5539E+04	0.1097E+03	50.48	4
0	0	13	3.173	0.2144E+01	0.2071E+02	0.10	4
0	0	14	2.946	0.2717E+04	0.6252E+02	43.46	4
0	0	15	2.750	0.1350E+02	0.2482E+02	0.54	4
0	0	16	2.578	0.1178E+04	0.4383E+02	26.88	4
0	0	17	2.426	-0.7420E+01	0.3549E+02	-0.21	4
0	0	18	2.292	0.4104E+03	0.4681E+02	8.77	4

and realize that this also tells us that the spacegroup is 77, not 75.

After his comes the table that tells us the quality of our data:

NOTE: Friedel pairs are treated as different reflections.

SUBSET OF INTENSITY DATA WITH SIGNAL/NOISE >= -3.0 AS FUNCTION OF RESOLUTION

RESOLUTION LIMIT	NUMBER OF REFLECTIONS OBSERVED	UNIQUE	POSSIBLE	COMPLETENESS OF DATA	R-FACTOR observed	R-FACTOR COMPARED expected	I/SIGMA	R-meas	Rmrgd-F
6.06	4189	556	560	99.3%	2.4%	2.7%	4187	66.74	2.6%	1.1%
4.31	7575	1008	1008	100.0%	2.6%	2.9%	7575	62.90	2.8%	1.2%
3.53	9468	1283	1283	100.0%	3.4%	3.2%	9468	53.37	3.6%	1.7%
3.06	11364	1540	1540	100.0%	5.1%	4.7%	11364	34.45	5.5%	3.1%
2.74	12628	1695	1695	100.0%	10.2%	10.4%	12628	17.09	11.0%	7.9%
2.50	14121	1916	1916	100.0%	21.5%	23.1%	14121	8.42	23.1%	17.1%
2.31	15155	2079	2079	100.0%	46.6%	50.5%	15155	3.92	50.2%	38.6%
2.16	12185	2104	2228	94.4%	113.3%	117.0%	12178	1.44	124.7%	119.0%
2.04	5134	1601	2347	68.2%	274.7%	291.2%	4913	0.40	325.5%	400.7%
total	91819	13782	14656	94.0%	5.7%	5.9%	91589	20.24	6.2%	15.0%

NUMBER OF REFLECTIONS IN SELECTED SUBSET OF IMAGES 93217
NUMBER OF REJECTED MISFITS 1391
NUMBER OF SYSTEMATIC ABSENT REFLECTIONS 0
NUMBER OF ACCEPTED OBSERVATIONS 91826
NUMBER OF UNIQUE ACCEPTED REFLECTIONS 13784

So the anomalous signal goes to about 3.3 Å (which is where 30% would be, in the "Anomal Corr" column), and the useful resolution goes to 2.16 Å, I'd say (pls note that this table treats Friedels separately; merging them increases I/sigma by another factor of 1.41).

For the sake of comparability, from now on we use the same axes (53.03 53.03 40.97) as the deposited PDB id 2QVO.

We could now modify XDS.INP to have

```
JOB=CORRECT  ! not XYCORR INIT COLSPOT IDXREF DEFPIX INTEGRATE CORRECT
SPACE_GROUP_NUMBER=   77
UNIT_CELL_CONSTANTS=    53.03   53.03  40.97  90.000  90.000  90.000
```

and run xds again, to obtain the final CORRECT.LP and XDS_ASCII.HKL with the correct spacegroup, but the statistics in 75 and 77 are the same, for all practical purposes (the 8 reflections known to be extinct do not make much difference).

Following this, we create XDSCONV.INP with the lines

```
SPACE_GROUP_NUMBER=   77  ! can leave out if CORRECT already ran in #77
UNIT_CELL_CONSTANTS=  53.03   53.03  40.97 90 90 90 ! same here
INPUT_FILE=XDS_ASCII.HKL
OUTPUT_FILE=temp.hkl CCP4
```

and run "xdsconv", and then

```
f2mtz HKLOUT temp.mtz<F2MTZ.INP
cad HKLIN1 temp.mtz HKLOUT output_file_name.mtz<<EOF
LABIN FILE 1 ALL
END
EOF
```

which gives us output_file_name.mtz, which we rename to xds-2ovo-1-F.mtz. Similarly, using

```
OUTPUT_FILE=temp.hkl CCP4_I
```

we end up with a MTZ file with intensities, which we rename to xds-2ovo-1-I.mtz.

dataset 2

This works exactly the same way as dataset 1. The geometry refinement is surprisingly bad:

```
REFINED PARAMETERS:  DISTANCE BEAM ORIENTATION CELL AXIS
USING   49218 INDEXED SPOTS
STANDARD DEVIATION OF SPOT    POSITION (PIXELS)     1.78
STANDARD DEVIATION OF SPINDLE POSITION (DEGREES)    0.15
CRYSTAL MOSAICITY (DEGREES)     0.218
DIRECT BEAM COORDINATES (REC. ANGSTROEM)   0.002198 -0.000174  0.526311
DETECTOR COORDINATES (PIXELS) OF DIRECT BEAM    1991.28   2027.42
DETECTOR ORIGIN (PIXELS) AT                     1984.09   2027.99
CRYSTAL TO DETECTOR DISTANCE (mm)       126.03
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  0.999979  0.002580 -0.006016
COORDINATES OF UNIT CELL A-AXIS   -31.728    -7.177   -42.595
COORDINATES OF UNIT CELL B-AXIS    40.575    13.173   -32.443
COORDINATES OF UNIT CELL C-AXIS    11.394   -39.576    -1.819
REC. CELL PARAMETERS   0.018658  0.018658  0.024258  90.000  90.000  90.000
UNIT CELL PARAMETERS     53.595    53.595    41.224  90.000  90.000  90.000
E.S.D. OF CELL PARAMETERS  1.0E-02 1.0E-02 1.7E-02 0.0E+00 0.0E+00 0.0E+00
SPACE GROUP NUMBER     75
```

with its large "STANDARD DEVIATION OF SPOT POSITION (PIXELS)" which may indicate a slipping crystal, or changing cell parameters due to radiation damage. However no indication of any of this is found in the repeated refinements listed in INTEGRATE.LP, so we do not know what to attribute this problem to!

The main table in CORRECT.LP is

```
     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-F
  LIMIT     OBSERVED  UNIQUE  POSSIBLE     OF DATA   observed  expected

    6.06        3925     547       560       97.7%       3.0%      3.3%     3922   56.13     3.3%     1.4%
    4.31        7498    1000      1000      100.0%       2.8%      3.4%     7498   56.91     3.0%     1.2%
    3.53        9407    1291      1291      100.0%       3.4%      3.5%     9407   52.39     3.7%     1.6%
    3.06       11005    1526      1526      100.0%       4.1%      3.9%    11005   42.13     4.4%     2.2%
    2.74       12569    1701      1701      100.0%       5.7%      5.7%    12569   28.38     6.1%     3.7%
    2.50       14020    1904      1904      100.0%       9.0%      9.9%    14020   17.92     9.7%     6.3%
    2.31       15101    2081      2081      100.0%      17.0%     19.0%    15101    9.83    18.3%    12.7%
    2.16       11693    2080      2202       94.5%      39.4%     40.8%    11682    4.00    43.6%    45.8%
    2.04        5152    1607      2345       68.5%      85.6%     93.5%     4943    1.21   101.3%   129.6%
   total       90370   13737     14610       94.0%       4.2%      4.5%    90147   24.22     4.6%     7.3%


NUMBER OF REFLECTIONS IN SELECTED SUBSET OF IMAGES   92690
NUMBER OF REJECTED MISFITS                            2318
NUMBER OF SYSTEMATIC ABSENT REFLECTIONS                  0
NUMBER OF ACCEPTED OBSERVATIONS                      90372
NUMBER OF UNIQUE ACCEPTED REFLECTIONS                13738
```

Dataset 2 is definitively better than dataset 1. Note that the number of misfits is more than 2.5% whereas one should expect about 1% (with WFAC1=1).

SHELXC/D/E structure solution

This is done in a subdirectory of the XDS data reduction directory (of dataset "1" or "2"). Here, we use a script to generate XDSCONV.INP (I used MERGE=TRUE, sometimes the results are better that way), run xdsconv and SHELXC.

```
#!/bin/csh -f

cat > XDSCONV.INP <<end
INPUT_FILE=../XDS_ASCII.HKL
OUTPUT_FILE=temp.hkl SHELX
MERGE=TRUE
FRIEDEL'S_LAW=FALSE
end

xdsconv

shelxc j <<end
SAD   temp.hkl
CELL 53.03 53.03 40.97 90 90 90
SPAG P42
MAXM 2
end
```

This writes j.hkl, j_fa.hkl and j_fa.ins. However, we overwrite j_fa.ins now (these lines are just the ones that hkl2map would write):

```
cat > j_fa.ins <<end
TITL j_fa.ins SAD in P42
CELL  0.98000  53.03   53.03  40.97   90.00   90.00   90.00
LATT  -1
SYMM -Y, X, 1/2+Z
SYMM -X, -Y, Z
SYMM Y, -X, 1/2+Z
SFAC S
UNIT   128
SHEL 999 3.0
FIND 3
NTRY 100
MIND -1.0 2.2
ESEL 1.3
TEST 0 99
SEED 1
PATS
HKLF 3
END
end
```

and then

```
shelxd j_fa
```

This gives best CC All/Weak of 37.28 / 21.38 for dataset 1, and best CC All/Weak of 37.89 / 23.80 for dataset 2.

Next we run G. Sheldrick's beta-Version of SHELXE Version 2011/1:

```
shelxe.beta j j_fa -a -q -h -s0.55 -m20 -b
```

and the inverse hand:

```
shelxe.beta j j_fa -a -q -h -s0.55 -m20 -b -i
```

One of these (and it's impossible to predict which one!) solves the structure, the other gives bad statistics.

Some important lines in the output: for dataset 1, I get

```
  78 residues left after pruning, divided into chains as follows:
A:  78

CC for partial structure against native data =  36.54 %

...

Estimated mean FOM and mapCC as a function of resolution
d    inf - 4.49 - 3.55 - 3.10 - 2.81 - 2.61 - 2.45 - 2.32 - 2.22 - 2.13 - 2.03
<FOM>   0.763  0.784  0.743  0.682  0.632  0.620  0.621  0.600  0.519  0.416
<mapCC> 0.890  0.936  0.916  0.893  0.838  0.827  0.847  0.858  0.836  0.768
N         721    728    722    720    719    738    749    721    674    721

Estimated mean FOM = 0.639   Pseudo-free CC = 65.26 %

Density (in map sigma units) at input heavy atom sites

 Site     x        y        z     occ*Z    density
   1   0.0293   0.3394   0.3145  16.0000    19.09
   2  -0.1598   0.3789   0.3723  12.7456    15.78
   3  -0.1413   0.4707   0.3704   9.4720     7.85
   4  -0.2238   0.1590   0.4520   9.2176     9.96
   5   0.0387   0.4228   0.3134   1.6608     1.28

Site    x       y       z  h(sig) near old  near new
  1  0.0293  0.3392  0.3148  19.1  1/0.02  2/10.34 4/11.66 4/11.66 5/12.88
  2 -0.1564  0.3740  0.3757  16.4  2/0.35  5/4.38 4/5.45 1/10.34 3/12.03
  3 -0.2146  0.1625  0.4495  11.0  4/0.53  2/12.03 5/15.84 1/16.92 4/17.39
  4 -0.1386  0.4748  0.3671   8.1  3/0.29  5/2.67 2/5.45 1/11.66 1/11.66
  5 -0.1829  0.4512  0.3605   5.9  3/2.47  4/2.67 2/4.38 1/12.88 1/13.92
```

and for dataset 2,

```
   80 residues left after pruning, divided into chains as follows:
A:  80

...

CC for partial structure against native data =  46.31 %
Estimated mean FOM and mapCC as a function of resolution
d    inf - 4.49 - 3.55 - 3.10 - 2.81 - 2.61 - 2.45 - 2.32 - 2.22 - 2.13 - 2.02
<FOM>   0.726  0.703  0.695  0.704  0.706  0.713  0.667  0.572  0.535  0.503
<mapCC> 0.850  0.863  0.857  0.899  0.900  0.908  0.866  0.805  0.828  0.814
N         719    721    725    719    713    736    755    722    673    705

Estimated mean FOM = 0.654   Pseudo-free CC = 67.40 %

Density (in map sigma units) at input heavy atom sites

 Site     x        y        z     occ*Z    density
   1   0.1613   0.5298   0.4706  16.0000    22.30
   2   0.1266   0.3414   0.5281  14.4576    17.03
   3   0.3453   0.2833   0.6078  11.1760    11.69
   4   0.0318   0.3665   0.5267   6.6512     8.45
   5   0.0499   0.3350   0.5280   5.8208     5.38

Site    x       y       z  h(sig) near old  near new
  1  0.1605  0.5316  0.4699  22.4  1/0.11  2/10.61 4/11.62 4/11.62 5/12.61
  2  0.1258  0.3407  0.5328  17.4  2/0.20  5/3.83 4/5.39 1/10.61 3/12.02
  3  0.3367  0.2831  0.6107  13.2  3/0.47  2/12.02 5/15.41 1/17.15 4/17.33
  4  0.0269  0.3630  0.5241   9.3  4/0.33  5/2.78 2/5.39 1/11.62 1/11.62
  5  0.0575  0.3206  0.5182   8.2  5/0.95  4/2.78 2/3.83 1/12.61 1/14.10
```

clearly indicating that the structure can be solved with each of the two datasets individually.

Can we do better?

data reduction

The safest way to optimize the data reduction is to look at external quality indicators. Internal R-factors, and even the correlation coefficient of the anomalous signal are of comparatively little value. A readily available external quality indicator is CC All/CC Weak as obtained by SHELXD, and the percentage of successful trials.

I tried a number of possibilities:

- Optimization by "re-cycling" GXPARM.XDS to XPARM.XDS and re-running INTEGRATE, coupled with REFINE(INTEGRATE)= ! (empty list) and specifying BEAM_DIVERGENCE_E.S.D. and similar parameters as obtained from INTEGRATE.LP: this quite often helps to improve geometry a bit but had no clear effect here.
- STRICT_ABSORPTION_CORRECTION=TRUE - this is useful if the chi^2 -values of the three scaling steps in CORRECT.LP are 1.5 and higher which is not the case here. Consequently this also had no clear effect.
- increasing MAXIMUM_ERROR_OF_SPOT_POSITION from its default of 3 to (3 * STANDARD DEVIATION OF SPOT POSITION (PIXELS)) which would mean increasing to 5 here: no clear effect.
- increasing WFAC1 : this was suggested by the number of misfits which is clearly higher than the usual 1 % of observations. WFAC1=1.5 has indeed a very positive effect on SHELXD: for dataset 1, the best CC All/Weak becomes **44.93 / 22.82** (dataset 2: **48.11 / 27.78**), and the number of successful trials goes from about 60% to 91% (dataset 2: 94%). **One should note that all internal quality indicators get worse when increasing WFAC1 - but the external ones got significant better!** The number of misfits with WFAC1=1.5 dropped to 196 / 436 for datasets 1 and 2, respectively.
- MERGE=FALSE vs MERGE=TRUE in XDSCONV.INP: after finding out about WFAC1 I tried MERGE=FALSE (the default !) and it turned out to be a bit better - best CC All/Weak **48.66 / 28.05** for dataset 2. On the other hand, the number of successful trials went down to 77% (from 94%). This result is somewhat difficult to interpret, but I like MERGE=TRUE better.

We may thus conclude that in this case the rejection of misfits beyond the target value of 1% reduces data quality significantly. In (other) desperate cases, if no successful trials are made by SHELXD it may be worth to always try WFAC1=1.5 provided the number of misfits is high.

We also learn that it's usually *not* going to help much to deviate from the defaults (MERGE=, MAXIMUM_ERROR_OF_SPOT_POSITION=, STRICT_ABSORPTION_CORRECTION=) unless there is a clear reason (high number of misfits) to!

structure solution

The resolution limit for SHELXD could be varied. For SHELXE, the solvent content could be varied, and the number of autobuilding cycles, and probably also the high resolution cutoff. Furthermore, it would be advantageous to "re-cycle" the file j.hat to j_fa.res, since the heavy-atom sites from SHELXE are more accurate than those from SHELXD, as the phases derived from the poly-Ala traces are quite good (compare the density columns of the two consecutive heavy-atom lists!).

With the optimally-reduced dataset 2, I get from SHELXE:

```
Density (in map sigma units) at input heavy atom sites

 Site     x         y         z      occ*Z     density
   1   0.3361   0.9695   0.9827  16.0000    24.15
   2   0.3708   1.1540   1.0380  14.5216    17.48
   3   0.1576   1.2210   1.1222   9.2848    12.60
   4   0.4807   1.1304   1.0314   7.2224     8.95
   5   0.4539   1.1750   1.0368   6.6224     7.26

Site    x       y       z  h(sig) near old  near new
  1  0.3380  0.9687  0.9828  24.3  1/0.11  6/2.40 2/10.33 4/11.42 4/11.81
  2  0.3732  1.1546  1.0426  18.1  2/0.23  5/4.00 4/5.67 6/9.92 1/10.33
  3  0.1637  1.2180  1.1226  13.5  3/0.36  2/12.06 5/15.47 6/15.97 1/17.12
  4  0.4784  1.1371  1.0333   9.3  4/0.38  5/2.89 2/5.67 1/11.42 1/11.81
  5  0.4439  1.1791  1.0300   9.0  5/0.64  4/2.89 2/4.00 6/12.54 1/12.64
  6  0.3273  0.9734  1.0393  -5.9  1/2.38  1/2.40 2/9.92 4/11.82 4/11.86
```

so the density is better, but not much. Furthermore, we note in passing that the number of anomalous scatterers (5) matches the sum of 4 Met and 1 Cys in the sequence.

Exploring the limits

With dataset 2, I tried to use the first 270 frames and could indeed solve the structure using the above SHELXC/D/E approach (with WFAC1=1.5) - 85 residues in a single chain, with "CC for partial structure against native data = 47.51 %". It should be mentioned that I also tried this in November 2009, and it didn't work with the version of XDS available then!

With 180 frames, it was possible to get a complete model by (twice) re-cycling the j.hat file to j_fa.res. **This means that the structure can be automatically solved just from the first 180 frames of dataset 2!**

Availability

- [1] (ftp://turn5.biologie.uni-konstanz.de/pub/xds-datared/2qvo/xds-2qvo-1-1_360-F.mtz) - amplitudes for frames 1-360 of dataset 1.
- [2] (ftp://turn5.biologie.uni-konstanz.de/pub/xds-datared/2qvo/xds-2qvo-1-1_360-I.mtz) - intensities for frames 1-360 of dataset 1.
- [3] (ftp://turn5.biologie.uni-konstanz.de/pub/xds-datared/2qvo/xds-2qvo-2-1_180-F.mtz) - amplitudes for frames 1-180 of dataset 2.
- [4] (ftp://turn5.biologie.uni-konstanz.de/pub/xds-datared/2qvo/xds-2qvo-2-1_180-I.mtz) - intensities for frames 1-180 of dataset 2.
- [5] (ftp://turn5.biologie.uni-konstanz.de/pub/xds-datared/2qvo/xds-2qvo-2-1_360-F.mtz) - amplitudes for frames 1-360 of dataset 2.
- [6] (ftp://turn5.biologie.uni-konstanz.de/pub/xds-datared/2qvo/xds-2qvo-2-1_360-I.mtz) - intensities for frames 1-360 of dataset 2.