S-SAD example

This example uses highly redundant data collected to a resolution of 2.1 Å at beamline BM14 of the ESRF at a wavelength of 1.722 Å. The protein contains 9 methionines and no cysteines. A high resolution dataset was also collected at ESRF to a resolution of 1.45 Å.

In this tutorial you will use SHELXC to prepare and analyze the data, SHELXD to find the sulphurs (and any other anomalous scatterers) and SHELXE to carry out density modification to break the phase ambiguity and to give phases and figures of merit to allow a map to be calculated.

The sequence of the protein is:

```
         10          20          30          40          50          60
MSAFDEFNEG FGLDVSDTPE ELAFETESAI EEIESETSPG DQPKGSEPEE IRVWAEEKAR

         70          80          90         100         110         120
KAVEEGREVT NWADWIMGWR TPNASEKKME FMYWYTRTYL EEAKDIRPDI ADALARGMAG

        130         140         150         160         170         180
LAFGRTDWVA SMLDPQIMRH IYTDPEVARI YSETRDMLRR VSDYYISLTT MELGKVADII

        190         200         210         220         230
AEAKAKGENP EVVAREIAEA VPRLSPKSLY FNLYYIGRSI GDNYVLEVAR VLSKMRRR
```

238 residues (27.3kDa) including 9 methionines (excluding start Met), no cysteines.

Data collected on BM14 at λ = 1.722 Å.

$P2_12_12_1$. a=42.89 b=53.71 c=95.20 Å. Estimated solvent content 40% with one monomer in the a.u.

Phasing data to 2.1 Å. 400 degrees at kappa=0, 360 degrees at kappa=30: **ssad.sca**

Statistics from scalepack (HKL200):

Shell Lower limit	Upper Angstrom	Average I	Average error	Norm. stat.	Linear Chi**2	Square R-fac	R-fac
30.00	4.52	315.5	2.6	1.7	1.284	0.036	0.040
4.52	3.59	350.0	4.1	2.1	1.253	0.050	0.060
3.59	3.14	212.3	2.8	1.6	1.369	0.061	0.068
3.14	2.85	126.1	1.6	1.4	1.340	0.064	0.067
2.85	2.65	83.0	1.4	1.2	1.171	0.079	0.079
2.65	2.49	64.7	1.3	1.2	1.118	0.097	0.092
2.49	2.37	55.2	1.3	1.3	1.112	0.115	0.110
2.37	2.26	44.1	1.4	1.3	1.122	0.144	0.140
2.26	2.18	35.1	1.4	1.4	1.119	0.188	0.174
2.18	2.10	26.6	1.5	1.4	1.184	0.252	0.224
All reflections		135.0	2.0	1.5	1.209	0.067	0.060

Shell Lower limit	Upper limit	Average Redundancy Per Shell
30.00	4.52	28.0
4.52	3.59	29.8
3.59	3.14	29.7
3.14	2.85	30.0
2.85	2.65	29.8
2.65	2.49	29.9
2.49	2.37	29.9
2.37	2.26	29.7
2.26	2.18	29.6
2.18	2.10	28.4
All hkl		29.5

High resolution data to 1.45 Å collected at λ = 0.9762 Å. **hires.sca**

Shell limit	Lower Angstrom	Upper	Average I	Average error	Norm. stat.	Linear Chi**2	Square R-fac	R-fac
30.00	3.12		1873.5	45.9	17.6	1.062	0.027	0.032
3.12	2.48		650.6	22.4	8.8	1.074	0.046	0.055
2.48	2.17		381.0	10.8	7.4	1.056	0.041	0.042
2.17	1.97		220.8	8.0	6.7	1.061	0.058	0.054
1.97	1.83		116.1	6.2	5.9	1.038	0.089	0.083
1.83	1.72		63.9	5.4	5.3	1.036	0.143	0.130
1.72	1.63		41.5	5.3	5.2	1.053	0.213	0.197
1.63	1.56		29.2	5.7	5.6	1.091	0.296	0.273
1.56	1.50		20.7	7.1	7.1	1.091	0.378	0.348
1.50	1.45		14.9	8.3	8.3	1.067	0.460	0.414
All reflections			367.4	13.0	7.8	1.060	0.045	0.037

You could use hkl2map to run SHELXC/D/E and visualise the statistics, or you could create a script using an editor (e.g. call it **psv**, enter the lines below, save it, type **chmod ugo+x psv**, and then enter **./psv** to run it). The key numbers are NTRY, the number of trials, and FIND, which should be within about 20% of the actual number of anomalous scatterers.

```
shelxc psv  << EOF
NAT hires.sca
SAD ssad.sca
CELL 42.89 53.71 95.20 90 90 90
SPAG P212121
FIND 9
SFAC S
NTRY 100
EOF
```

This will create psv_fa.ins (input for SHELXD), psv_fa.hkl (“heavy atom” structure factors h,k,l,FA, sigFA, α) and psv.hkl (high resolution structure factors, h,k,l,F,sigF)

To run SHELXD, type

shelxd psv_fa

this tries to find the anomalous scatterers. The key numbers are CC All/Weak. These should be greater than 30/15 for a meaningful solution for SAD. This produces psv_fa.res (heavy atom coordinates in shelx format), psv_fa.pdb (heavy atom coordinates in pdb format) and psv_fa.lst (detailed output from SHELXD). Note how the occupancies of the heavy atoms vary.

You can now use SHELXE to use these heavy atom positions to calculate phases, with density modification being used to break the phase ambiguity. You need to try both hands of the heavy atoms. The –h switch indicates that the heavy atoms are in the native data (i.e. psv.hkl), the –b switch creates a file of the refined heavy atom positions, the –s0.4 indicates a 40% solvent content, -m20 indicates 20 cycles of density modification and –i tells shelxe to change the hand of the heavy atoms.

shelxe psv psv_fa –h –b –s0.4 –m20

shelxe psv psv_fa –h –b –s0.4 –m20 -i

the key numbers are the Contrast and the Pseudo-free CC. There should be a clear discrimination between the two runs to indicate the correct hand. The files produced are psv.phs (the protein phases, file containing h,k,l,F,sigF,phi,fom), psv.pha (the anomalous scatter phases), psv.hat (refined heavy atom positions – this can be read into coot), and psv.lst (copy of the output of shelxe). A similar set of files is created for the second run: psv_i.phs, psv_i.pha, psv_i.hat and psv_i.lst.

You can now coot to look at the maps calculated from the phases. Read in the appropriate .hat file first, and then the .phs file.

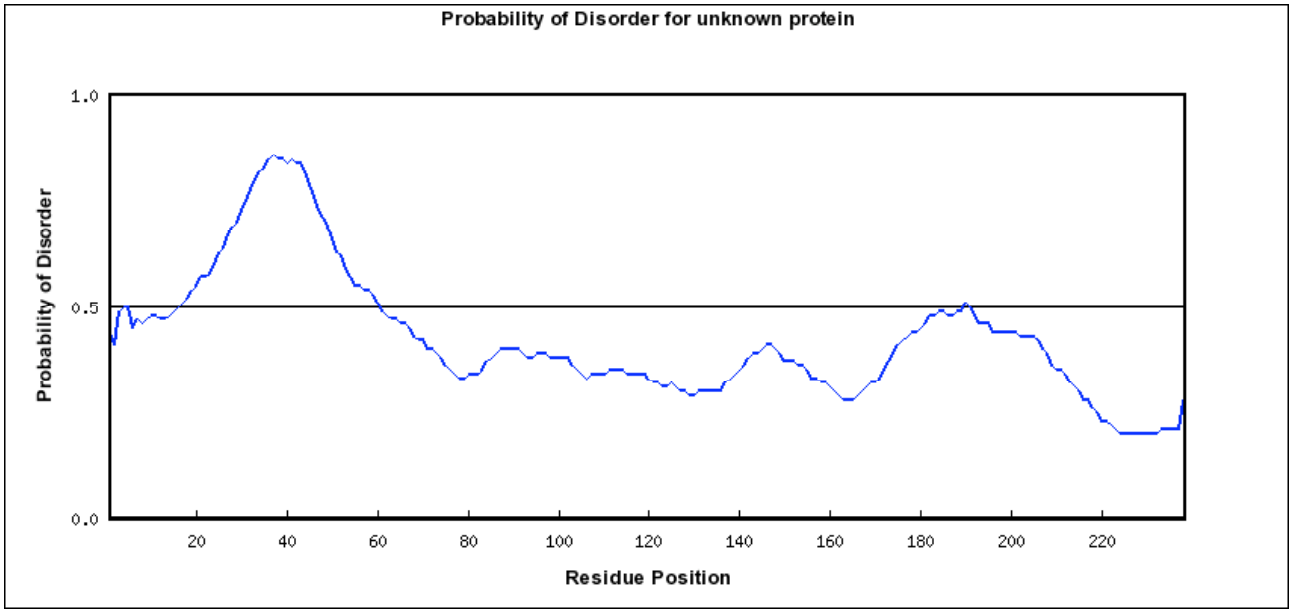
Is the map interpretable ?

You could try running SHELXE again with a higher solvent content. Sometimes this helps.

As the high resolution data extend to 1.45 Å, you could try the “free lunch” switch. Just do this on the correct solution by adding –e1.0 to the SHELXE command line.

Garry Taylor, University of St Andrews (glt2@st-andrews.ac.uk)

Data courtesy of the Scottish Structural Proteomics Facility.


Probability of Disorder for unknown protein
Probability of Disorder
1.0
0.5
0.0
20
40
60
80
100
120
140
160
180
200
220
Residue Position