MAD on Alba

This example uses data collected to 2.7 Å at 3 wavelengths on ID14-3 at the ESRF. The protein contains two selenomethionines per monomer, with two monomers estimated to be in the asymmetric unit. There is also a native dataset (no seleniums) to 2.6 Å.

In this tutorial you will use SOLVE/RESOLVE and/or SHELX to solve the structure. The sequence of the protein is:

```
         10          20          30          40          50          60
MEKMSSGTPT PSNVVLIGKK PVMNYVLAAL TLLNQGVSEI VIKARGRAIS KAVDTVEIVR

         70          80          90         100
NRFLPDKIEI KEIRVGSQVV TSQDGRQSRV STIEIAIRKK
```

99 residues (10.8kDa) including 2 selenomethionines (excluding start Met).

Three wavelength data collected, with estimates of f′ and f′′ from the fluorescence scan of the crystal on the beamline:

Peak (λ1)	at λ = 0.979 Å. f′ = -6.2, f′′ = 6.0	**peak.sca**
Inflection (λ2)	at λ = 0.980 Å. f′ = -10.2, f′′ = 3.06	**inf.sca**
Remote (λ3)	at λ = 0.939 Å. f′ = -3.02, f′′ = 4.40	**rem.sca**

Native data is **nat.sca**. $P6_122$ or $P6_522$. a=b=84.43 c=162.24 Å. Estimated solvent content 70% with two monomers in the a.u.

You can try solving this using SHELXC/D/E and SOLVE/RESOLVE.

SHELX. 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 **alba**, enter the lines below, save it, type **chmod ugo+x alba**, and then enter **./alba** 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  alba << EOF
NAT  nat.sca
HREM rem.sca
PEAK peak.sca
INFL inf.sca
CELL 84.43 84.43 162.24 90 90 120
SPAG P6522
FIND 4
NTRY 100
EOF
```

This will create alba_fa.ins (input for SHELXD), alba_fa.hkl ("heavy atom" structure factors h,k,l,FA, sigFA, α) and alba.hkl (high resolution structure factors from nat.sca, h,k,l,F,sigF)

To run SHELXD, type

shelxd alba_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 – but it may be less. This produces alba_fa.res (heavy atom coordinates in shelx format), alba_fa.pdb (heavy atom coordinates in pdb format) and alba_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 –b switch creates a file of the refined heavy atom positions, the –s0.7 indicates a 70% solvent content, -m20 indicates 20 cycles of density modification and –i tells shelxe to change the hand of the heavy atoms.

shelxe alba alba_fa –b –s0.7 –m20

shelxe alba alba_fa –b –s0.7 –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 alba.phs (the protein phases, file containing h,k,l,F,sigF,phi,fom), alba.pha (the anomalous scatter phases), alba.hat (refined heavy atom positions – this can be read into coot), and alba.lst (copy of the output of shelxe). A similar set of files is created for the second run: alba_i.phs, alba_i.pha, alba_i.hat and alba_i.lst.

You can now use **coot** to look at the maps calculated from the phases. Read in the appropriate .hat file first, and then the .phs file. Also read in the .pha file to see the anomalous difference Fourier map – this clearly locates the anomalous scatterers. Try skeletonizing the map in coot, this helps you see any secondary structure.

Is the map interpretable ?

You could try running SHELXE again, this time just for the correct hand and try varying the solvent content and/or number of cycles of density modification. Sometimes this helps.

No free lunch here, as the data only extend to 2.6 Å!!

The refined structure is in the file **ss.pdb** if you want to see how good the density is.

You could use **ccp4i** to convert the .phs file to a .mtz file, and then run **Arp/Warp** to see if it can interpret the map. You can read the .phs file into CCP4i using "Convert to/modify/extend MTZ" with a user defined format of (3f4.0,f9.2,f8.4,f8.1,f8.4) for H,K,L, FP, FOM, PHI, SIGFP. This can then be used to input to Arp/Warp. The sequence of the protein for Arp/Warp is in the file **seq.pir**.

SOLVE/RESOLVE

Use the script **mad.com** to run SOLVE/RESOLVE.

The asymmetric unit contains two monomers. Once part of the structure is built, the NCS can be defined, and RESOLVE run again using NCS-averaging. **resolve.com** contains the NCS and can be run – see if the map is any better. Resolve can derive the NCS from the heavy atom positions, however in this case with only one Se per monmer, it is impossible to do this. You need at least 3 atoms in each monomer.