A difficult SAD case with SHELXC/D/E

Andrea Thorn

Overview

The prerequisite for this tutorial is the basic SHELXC/D/E tutorial.
In this tutorial, we will solve a relatively difficult SAD case, fibronectin.

Molecule:	Fibronectin
Number of animo acids	90
Resolution	1.51 Å
Space group	$P4_12_12$ or $P4_32_12$
Cell	a=b=37.86Å c=108.42Å
	α=90° β=90° γ=90°

Rudino-Pinera, Ravelli, Sheldrick, Nanao, Korostelev, Werner, Schwarz-Linek, Potts & Garman, JMB 368 (2007), 833-844.

Data preparation with SHELXC

We start with two files: fibro-nat.sca contains high-resolution data; fibro-sad.sca contains long-wavelength high-multiplicity data but to lower resolution. The input file for SHELXC should use all available data, e.g. both files. The native high-resolution data will be useful in SHELXE density modification. The protein contains 4 disulfide bridges, 2 methionine residues and the buffer contained sulfate. We do not know how many sulfates may have bound to the protein. In addition, we do not know whether the correct space group is $P4_12_12$ or $P4_32_12$. For the marker atom search, it will not play a role if the symmetry is inverted, as it does not matter for the substructure itself. SHELXE, if run with the –i option, can be used to discern between the hand, and if necessary, will put the symmetry operators and the space group to the correct enentomorph. So, you do not have to worry and can just use $P4_32_12$.

`CELL 37.86 37.86 108.42 90 90 90`	Cell parameters
`SPAG P43212`	Space group
`FIND 10`	Number of marker atoms: 8 CYS + 2 MET
`NTRY 100`	Number of tries (100 is here sufficient)
`SAD fibro-sad.sca`	Filename of data set
`NAT fibro-nat.sca`	
`SFAC S`	Marker atom type; only for naming

We then start SHELXC:
`shelxc fibro < myinputfile`

The screen output gives:

```
Resl.   Inf - 8.0 - 6.0 - 5.0 - 4.0 - 3.5 - 3.0 - 2.8 - 2.6 - 2.4 - 2.2 - 2.00
 N(data)    117    141    175    366    353    625    386    511    685    954   1333
 Chi-sq    1.05   0.66   0.76   1.10   1.92   3.55   2.49   2.27   1.93   2.06   3.65
 <I/sig>  142.7  158.5  163.9  154.9  142.8  115.7   95.4   79.6   59.6   40.2   18.9
 %Complete 87.3   96.6   98.3   94.8   97.0   95.6   97.0   97.3   97.2   96.4   94.5
 <d"/sig>  7.45   6.85   4.49   3.31   2.63   2.03   2.05   1.80   1.57   1.26   1.07
 CC(1/2)   96.9   97.9   94.6   90.6   78.1   61.7   53.0   53.9   36.9   20.3   12.6
```

If `SHEL <lower limit> <higher limit>` has not been specified in the SHELXC input, the program automatically includes a resolution cut-off that is 0.5Å from the resolution limit of the data (`SHEL 999 2.5`). A better guide is where CC(1/2) drops below 25%, hence a cut-off between 2.4 and 2.2 would be also feasible. As disulfide bridges are 2.03 Å long and the resolution would be less than this, the number of peaks to search for (FIND) has to be changed in any case. We are searching for 4 disulfide bridges and 2 methionines instead of 8 cysteine and 2 methionine residues – hence you should replace "FIND 10" with "FIND 6" in either the SHELXC input (and run SHELXC again) or in SHELXD.

Output files from SHELXC are:

fibro_fa.hkl   contains $||F_{hkl}|-|F_{-h-k-l}||$ (Bijvoet differences) and estimates for the phase shifts α

fibro_fa.ins   contains instructions for SHELXD

fibro.hkl   contains intensity data for SHELXE

Alternative: Data preparation with XPREP

Start xprep by with the terminal command
xprep

Give the file name of the native data file (fibro-nat.sca). Choose the correct input file format (here: DENZO/SCALEPACK or HKL2000 .sca file). The cell is read in automatically from the .sca file, you just have to confirm to use this cell. Note that only merged .sca files contain the cell, so the native data must be already merged! Why could that be a problem for the anomalous data?

We do not yet know which space group enantiomorph we have. However, without having solved the structure, we do not have means to differentiate. SHELXD will run with both $P4_32_12$ and $P4_12_12$ ok, because the marker atom search is not influenced by handedness of the substructure or symmetry elements. SHELXE density modification will later allow us to differentiate.

Go quickly through the lattice and space group determination and just use one of the two possible choices: $P4_32_12$.

After the space group is established correctly - you can check in the header – you have to use the option '[D] Read, modify or merge DATASETS'.

Choose '[R] read in another dataset' and read in fibro_sad.sca as being indexed on the original cell.

```
 Select option [S]: r

 Input file indexed on original (O), current cell (C) or neither (N) [O]: o

Filename ( .hkl assumed): fibro-sad.sca

 [4] SHELX HKLF 4 format or SAINT ._rf file
 [3] SHELX HKLF 3 format
 [C] SHELX condensed data (HKLF -1)
 [N] XENGEN/XGEN .mu/.mui file (F-squared)
 [M] XENGEN MULIST .mu file (F)
 [D] DENZO/SCALEPACK or HKL2000 .sca file
 [X] XDS_ASCII.HKL format from XDS or XSCALE
 [E] EXIT to main menu
 [Q] QUIT program

 Select option [D]: d

 ** Data multiplied by   0.0100  to bring onto reasonable scale **

  221819 Reflections read from file fibro-sad.sca
   Mean (I/sigma) =   12.03
```

By using '[S] Display intensity STATISTICS' and then '[P] Rpim, Ranom etc. – NEEDS UNMERGED DATA', you could have a look at the data statistics. After you have become familiar with the data sets, choose '[A] MAD, SAD, SIR or SIRAS' and then '[A] SAD (Single-wavelength Anomalous Scattering)'. The target number of reflections in local scaling sphere can be the default 100 for almost all cases.

A table of the anomalous signal is displayed:

```
 [M] MAD (Multiple-wavelength Anomalous Dispersion)
 [I] SIR (Single Isomorphous Replacement)
 [A] SAD (Single-wavelength Anomalous Scattering)
 [R] SIRAS (Single Isomorphous Replacement with Anomalous Scattering)
 [E] EXIT to main menu
```

```
 [Q] QUIT program

 Select option [E]: a

 Target number of reflections in local scaling sphere
 (0 if no local scaling) [100]:

Anomalous signal/noise ratios (1.0 is random).  The first line is based on
 input sigmas, the second on variances of F+ and F- (if not already averaged):
 Inf - 8.0 - 6.0 - 5.0 - 4.0 - 3.5 - 3.0 - 2.7 - 2.5 - 2.3 - 2.1 - 1.9 - 1.7 A
    5.05  5.06  3.99  2.99  2.68  2.08  1.90  1.77  1.60  1.49  1.26  1.11
    4.95  4.99  3.95  3.03  2.59  2.11  1.89  1.70  1.57  1.40  1.14  1.04

   100.4  Neighbors used on average for F+/F- local scaling
 Ranom =   0.0450  before and   0.0451  after local scaling

 Enter effective B-value (e.g. 20) to normalize delta-F or Fa values,
 <CR> for no renormalization: 20
```

Then you should give the effective B value for normalization. You should in any case give a value. 20 is usually good. The program then requires a file name for the anomalous differences and α angle estimates. This filename should end with _fa.hkl to comply with the SHELXC standard. (For example, fibro_fa.hkl)

```
Filename to write FA and phi(T)-phi(A) (<CR> for none): fibro_fa

Current dataset contains  4224 SAD delta(F)
```

Choose to write an .ins file for SHELXD and give it the same first part of a file name as the previously written file.

```
 Write .ins file for SHELXD/XD (Y or N)? [Y]: Y

 Filename [fibro_fa.ins]: fibro_fa.ins
```

You should specify the atom type you expect, here this is sulfur. Disulfide bridges are 2.03 Å long, so if we cut the resolution to 2.1 or lower, we have to use the number of disulfides instead of single sulfur atoms. The disulfide peaks will fuse into one anomalous density peak! Hence, we give 6: 4 disulfide bridges and 2 methionine residues.

```
 Element type for heavy atoms [Se]: S

 Number of unique heavy atoms [12]: 6
```

As for the wavelength, it was 1.771 Å. For a SAD experiment, the wavelength and the specified atom type are not of importance! Give a resolution cut-off of 2.4 Å.

```
 Wavelength [1.54178]: 1.771

 Resolution cutoff in Angstroms [ 2.4 ]: 2.4
```

To be able to run SHELXE later, you should now use the option '[C] Change CURRENT dataset' to change back to the first data set and then write it out with the option '[W] WRITE dataset to file' as SHELX hkl file without R(free) flags and give it the same name: fibro.hkl. The R(free) flags are only useful if you intend to refine later with SHELXL. You can now exit XPREP, and proceed with SHELXD.

Finding the marker atom substructure with SHELXD

Run
shelxd fibro_fa
or
shelxd_mp fibro_fa

Output files from SHELXD are:
fibro_fa.res   marker atom solution with the best CFOM; this file is updated while SHELXC runs
fibro_fa.lst   log file

Repeat this with the other space group. The best CFOM-values are between 35% and 42%, which might be a solution, but not a good one – however, we can try to get a better CFOM by using DSUL in the .ins file (the instructions file for SHELXD), which will explicitly search for elongated peaks. Just include DSUL 4 to search for 4 disulfide bridges and make sure MIND is set to MIND -3.5. We will also heighten the number of tries by setting "NTRY 1500".
(If does not give you a significant higher CFOM value, you can also change the resolution limit by using, for example SHEL 999 2.4 in the .ins file.) This is an exemplary instructions file:

`TITL fibro_p43212_fa.ins SAD in P43212`	Title. Irrelevant to SHELX.
`CELL  0.98000 37.86 37.86 108.42 90.00 90.00 90.00`	Cell parameters
`LATT  -1`	Lattice type
`SYMM 1/2-Y, 1/2+X, 3/4+Z`	Symmetry operators
`SYMM -X, -Y, 1/2+Z`	
`SYMM 1/2+Y, 1/2-X, 1/4+Z`	
`SYMM 1/2-X, 1/2+Y, 3/4-Z`	
`SYMM Y, X, -Z`	
`SYMM 1/2+X, 1/2-Y, 1/4-Z`	
`SYMM -Y, -X, 1/2-Z`	
`SFAC S`	Marker atom type; only for naming
`UNIT  192`	Number of atoms in the unit cell. Irrelevant to SHELXC/D/E.
`SHEL 999 2.5`	Resolution cut-offs for SHELXD.
`PATS`	Use Patterson seeding.
`DSUL 4`	Find 4 disulfide bridges.
`FIND 6`	Find 6 anomalous density peaks (4 disulfide bridges, 2 MET residues).
`MIND -3.5`	The minimum distance between peaks is 3.5Å.
`NTRY 1500`	Do 1500 tries.
`SEED 1`	Use 1 as a seed (this ensures the same starting point if repeated).
`HKLF 3`	Input format is F_A values, hence HKLF 3. (HKLF 4 would be intensities.)
`END`	End of file.

You should now be able to get best CFOM values over 60, almost certainly a good solution.

Density modification and auto-tracing with SHELXE

First, we need to calculate the solvent content. The volume of the rectangular unit cell is approximately 38 Å x 38 Å x 108 Å = 155952 Å^3. There are 8 asymmetric units per unit cell, which can be looked up in the International Tables A. An average amino acid in a crystal occupies 140 Å^3 of space. We expect 90 amino acid residues in the asymmetric unit. They would hence occupy 90 x 140 Å^3 x 8 = 100800 Å^3. 100800 Å^3 divided by 155952 Å^3 is approximately 0.65, which can then be the protein content. The solvent content is hence 0.35, rather low and hence unfavorable for density modification.

Please use these two commands in two windows, but in the same folder:
shelxe fibro fibro_fa -a -h -s0.35

```
shelxe fibro fibro_fa -a –h –s0.4 -i
```

This will read in
fibro.hkl native data
fibro_fa.hkl anomalous differences, α
fibro_fa.res marker atom substructure

A CC of partial structure against native data over 25% indicates a correct solution (and consequently, the correct space group choice).

The output files are:
fibro.lst Log file
fibro.pha Anomalous density map
fibro.pdb Backbone trace
fibro.hat Improved substructure coordinates (hat stands for heavy atom solution)
fibro.phs hkl file with phases to display electron density
The files from the run with the inverted substructure will have "_i" attached, for example "fibro_i.pha".

If you find that the structure cannot be traced at your first try, you should try one of the following options (or a combination thereof):
- Use **–z** to optimize the marker atom substructure before density modification and auto-tracing
- Use **–q** to search for helices.
- Use more auto-tracing macrocycles (e.g. **–a20**)

If the structure is not immediately solved, this might be due to not all marker atom positions being correct, a high variance in B factors, the wrong hand is used or the structure is not containing any helices. (Pure beta sheet structures are more difficult to trace.)

After you have solved the structure, have a look at the output coordinates and maps in COOT!