

Basic SHELXC/D/E tutorial

Andrea Thorn

Overview

In this tutorial, SHELX will be used for experimental phasing. Two cases will be processed: An easy S-SAD case and a MAD case. Several additional tasks are given in the end. We will use the command line and COOT for this. I am very grateful to Tim Gruene and George Sheldrick for their kind advice!

An online version of this tutorial hand-out can be found at: <http://shelx.uni-ac.gwdg.de/~athorn>

Program availability

SHELX is available from <http://shelx.uni-ac.gwdg.de/SHELX/> and free for academic use.

SHELX beta versions can be obtained from Prof. George Sheldrick, gsheldr@shelx.uni-ac.gwdg.de

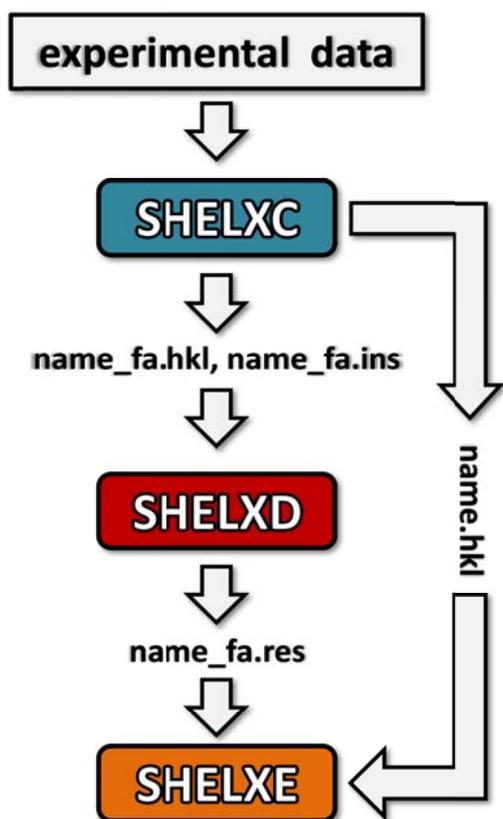
MAC versions are only compatible with an Intel CPU and 64-bit OS, not with PPC or 32 bit.

XPREP is a proprietary Bruker program, but not essential to run SHELXC/D/E. A free one-year trial version of XPREP can be obtained from Bruker.

Installation

To install SHELXC/D/E on your computer put the executable files into a directory in the 'PATH'.

General data flow



SHELXC (XPREP) sets up files and calculates the estimated substructure structure factor amplitude $|F_A|$ and the phase shift α from input data.

name_fa.ins cell, symmetry and instructions for SHELXD

name_fa.hkl $h, k, l, |F_A|, \sigma(|F_A|)$ and the estimated α angles

name.hkl $h, k, l, |F_{obs}|^2, \sigma(|F_{obs}|^2)$

SHELXD finds the marker atom substructure.

name_fa.res potential marker atom positions

SHELXE applies density modification to establish the correct handedness, if necessary, and can autotrace the poly-Ala backbone given good resolution native data.

name.phs contains information for map generation

name.hat improved marker atom positions

name.pdb if auto-tracing has been used

name.pha phases for marker atom substructure

S-SAD Phasing

Molecule:	Viscotoxin-A1
Number of amino acids	92
Resolution	1.25 Å
Space group	P4 ₃ 2 ₁ 2
Cell	a=b=65.73Å c=47.16Å α=90° β=90° γ=90°
Strategy	In-house CuK _α (λ=1.54178 Å) for S-SAD

Pal A, Debreczeni JE, Sevvana M, Gruene T, Kahle B, Zeeck A, Sheldrick GM. *Acta Crystallogr. D64 (2008), 985-92.*

These data were collected by Aritra Pal on a Cu home source. The solution of this structure is also discussed in Bernhard Rupp's book "Biomolecular Crystallography", Garland Science 2010. pp. 530-535.

Data preparation with SHELXC

We start with the data contained in the file `anom_data.hkl`. SHELXC can read data from XDS (`XDS_ASCII.HKL`), sca files and SHELX hkl files. It is best to read in `XDS_ASCII.HKL` directly rather than processing it with other programs, so that the data are *unmerged*, which is very desirable. You should create a new file, which could have any name. We choose 'visc-a1'. It should contain the following lines:

<code>CELL 65.73 65.73 47.16 90 90 90</code>	Cell parameters
<code>SPAG P43212</code>	Space group
<code>FIND 12</code>	Number of marker atoms*
<code>NTRY 100</code>	Number of tries (100 is here sufficient)
<code>SAD anom_data.hkl</code>	Filename of data set
<code>SFAC S</code>	Marker atom type; only for naming

* NOTE: Disulfide bridges are 2.03 Å long, so if we cut the resolution to 2.1 or less, we have to change this number: The disulfide peaks should then fuse into one anomalous density peak!

```
shelxc visc-a1 < visc-a1
```

Note that the first 'visc-a1' is the project name you want to assign and is used to generate the names of the files that are written by SHELXC. The second is the name of the input file. Having the same name makes it easier to find the input file later, but is not required!

Look at your screen output! There is a table like this:

Resl.	Inf	- 8.0	- 6.0	- 5.0	- 4.0	- 3.5	- 3.0	- 2.5	- 2.3	- 2.1	- 1.9	- 1.70
N(data)	148	173	216	475	458	813	1588	1054	1496	2152	3232	
Chi-sq	1.04	1.00	0.97	0.90	0.93	0.90	0.94	1.05	1.12	1.15	1.02	
<I/sig>	128.2	126.1	125.1	140.6	128.4	108.7	72.1	54.2	37.2	15.5	4.7	
%Complete	94.9	98.3	98.6	99.4	99.3	99.5	99.7	99.8	99.7	99.8	99.5	
<d"/sig>	3.83	3.61	2.96	2.18	1.95	1.57	1.33	1.23	1.11	0.88	0.75	
CC(1/2)	87.2	90.4	74.8	73.5	59.5	51.9	35.3	28.9	20.7	5.2	-4.4	

It gives against resolution:

- The number of reflections read in
- χ^2 which should be about 1 if the agreement of equivalents agrees with their esds.
- The average intensity divided by its standard deviation σ
- Completeness (in %)
- d" divided its estimated standard deviation σ , giving you a good indication for the strength of the anomalous signal, which should asymptote to 0.8 in the outer shell, if the data are processed well. If this line is missing, you should check if you gave an input data set in which the Friedel opposites have been merged.
- The self-correlation coefficient for the anomalous signal. It should be above 25% for a significant signal.

In SHELXD, the anomalous differences will be used to find the marker atom positions. Where should you cut the data for SHELXD? This value should be varied for more difficult cases, as it is often crucial!

Output files from SHELXC are:

visc-a1_fa.hkl contains $||F_{hkl}| - |F_{-h-k-l}||$ (Bijvoet differences) and estimates for the angles α
visc-a1_fa.ins contains instructions for SHELXD
visc-a1.hkl contains intensity data for SHELXE

You can look at these files with a text editor. Where is the resolution cut-off for SHELXD?

Alternative: Data preparation with XPREP

Start xprep by with the terminal command
xprep

Give the file name of the data file you want to use (anom_data.hkl). If, for example for MAD, there are several files, just give one of them. XPREP can read data in XDS (XDS_ASCII.HKL), sca and SHELX hkl formats. Choose the correct input file format (here: SHELX HKLF 4 format). Then you have to give the cell in the format 'a b c α β γ '.

```
478104 Reflections read from file anom_data.hkl
Mean (I/sigma) = 9.21
```

```
Enter cell (Angstroms and degrees) corresponding to indices in file (if any):
65.73 65.73 47.16 90 90 90
```

As we also know the space group, you can go rather quickly through lattice and space group determination by hitting 'Enter'. Why does the program propose both $P4_32_12$ and $P4_12_12$?

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'.

```
Current dataset: anom_data.hkl           Wavelength: 0.71073 Chiral: ?
-----
Original cell: 65.730 65.730 47.160 90.00 90.00 90.00 Vol 203751.6
Esds: 0.000 0.000 0.000 0.00 0.00 0.00 Lattice: P
-----
Current cell: 65.730 65.730 47.160 90.00 90.00 90.00 Vol 203751.6
-----
Matrix: 1.0000 0.0000 0.0000 0.0000 1.0000 0.0000 0.0000 0.0000 1.0000
-----
Crystal system: Tetragonal Space group: P4(3)2(1)2 # 96 [chi] Laue: 5
-----
```

```
[D] Read, modify or merge DATASETS           [C] Define unit-cell CONTENTS
[P] Contour PATTERSON sections               [F] Set up shelxtl FILES
[H] Search for HIGHER metric symmetry       [R] RECIPROCAL space displays
[S] Determine or input SPACE GROUP          [U] UNIT-CELL transformations
[A] Absorption, powder, SIR, SAD, MAD etc. [T] Change TOLERANCES
[M] Test for MEROHEDRAL TWINNING           [O] Self-rotation function
[L] Reset LATTICE type of original cell     [Q] QUIT program
```

```
Select option [D]: D
```

Choose '[S] Display intensity STATISTICS' and you can see your normal data statistics under option '[A] Merge all equivalents'.

```

Index  # Data  Filename or Source of Data

      1   478104   anom_data.hkl  <- current dataset

```

```

[M] Sort-MERGE current data (no scaling)      [C] Change CURRENT dataset
[L] LEAST-SQUARES scale and merge datasets    [W] WRITE dataset to file
[I] INCLUDE Rfree flags from another file    [R] READ in another dataset
[S] Display intensity STATISTICS             [D] DELETE stored dataset
[F] FACE-indexed absorption corrections      [P] PSI-scan absorption corr.
[T] Copy file, TRANSFORM hkl and cosines    [A] MAD, SAD, SIR or SIRAS
[H] Apply HIGH/low resolution cutoffs       [N] NORMALIZE/scale sigmas
[G] Generate simulated powder diagrams      [U] Anisotropic scaling
[Y] Self or cross correlation coefficients   [J] Change dataset label
[Z] Expand data to triclinic                [E] EXIT to main menu
[X] Parsons Q values and Flack x parameter  [Q] QUIT program

```

Select option [S]: **S**

These are also printed to the file anom_data.prp and can later be used for data tables. After that, you will be back in the datasets menu. Choose again '[S] Display intensity STATISTICS', but this time you should choose '[P] Rpim, Ranom etc. - NEEDS UNMERGED DATA'.

Current dataset: visc-a1_anomdata.hkl

```

[R] Define resolution ranges (currently selected automatically)
[N] Do NOT merge reflections
[I] Merge IDENTICAL indices only
[S] Merge SYMMETRY equivalents (but not Friedel opposites)
[A] Merge ALL equivalents (including Friedel opposites)
[P] Rpim, Ranom etc. - NEEDS UNMERGED DATA
[E] EXIT to data processing menu
[Q] QUIT program

```

Select option [A]: **P**

Resolution	#Data	#Theory	%Comp	%F-comp	Redund	<d"/sig>	Ranom	Rrim	Rpim	
Inf - 7.61	178	179	99.4	100.0	54.68	3.94	0.0285	0.0370	0.0046	
7.61 - 4.88	414	415	99.8	100.0	67.89	3.08	0.0333	0.0426	0.0051	
4.88 - 3.80	598	599	99.8	99.8	70.81	2.18	0.0219	0.0391	0.0046	
3.80 - 3.29	588	588	100.0	100.0	72.80	1.76	0.0209	0.0429	0.0050	
3.29 - 2.97	601	601	100.0	100.0	73.46	1.47	0.0211	0.0504	0.0058	
2.97 - 2.75	585	585	100.0	100.0	73.86	1.40	0.0269	0.0658	0.0076	
2.75 - 2.58	600	600	100.0	100.0	71.06	1.31	0.0327	0.0841	0.0099	
2.58 - 2.44	612	612	100.0	100.0	66.67	1.27	0.0351	0.0903	0.0110	
2.44 - 2.33	599	599	100.0	100.0	63.98	1.19	0.0345	0.0934	0.0116	
2.33 - 2.24	569	569	100.0	100.0	61.60	1.17	0.0433	0.1203	0.0152	
2.24 - 2.16	599	599	100.0	100.0	51.68	1.11	0.0467	0.1241	0.0170	
2.16 - 2.09	592	592	100.0	100.0	34.92	1.01	0.0610	0.1478	0.0246	
2.09 - 2.03	570	572	99.7	99.6	25.95	0.89	0.0693	0.1605	0.0309	
2.03 - 1.97	646	646	100.0	100.0	20.05	0.93	0.1040	0.2034	0.0444	
1.97 - 1.92	598	598	100.0	100.0	14.95	0.86	0.1323	0.2373	0.0597	
1.92 - 1.87	675	676	99.9	99.8	11.01	0.94	0.2075	0.2753	0.0804	
1.87 - 1.83	578	578	100.0	100.0	7.79	0.86	0.2501	0.3067	0.1067	
1.83 - 1.80	469	469	100.0	100.0	5.18	0.83	0.4013	0.3731	0.1615	
1.80 - 1.76	677	677	100.0	100.0	4.59	0.77	0.3392	0.3184	0.1473	
1.76 - 1.73	550	550	100.0	100.0	4.53	0.78	0.4055	0.3459	0.1606	
1.73 - 1.70	546	560	97.5	97.8	4.29	0.79	0.4288	0.3683	0.1738	

1.80 - 1.70	1773	1787	99.2	99.3	4.48	0.78	0.3843	0.3404	0.1586	
Inf - 1.70	11844	11863	99.8	99.8	40.30	1.19	0.0514	0.0633	0.0086	

Merged [P], lowest resolution = 47.16 Angstroms

This gives you a table of several indicators for the data quality and the anomalous signal. Of particular importance is the $\langle d''/\sigma \rangle$ value, giving you a good indication for the strength of the anomalous signal, which should asymptote 0.8 in the outer shell, if the data are processed well. Think were you would want to cut your data for the marker atom search.

> **Why does this need unmerged data?**

If you like, you can also generate plots, for example d''/σ against resolution. When you are done, you return to the datasets menu. Now you should choose option '[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
```

Please note that due to an older estimation procedure, the values in this table asymptote 1.0 for no signal. 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, `visc-a1_fa.hkl`)

```
Filename to write FA and phi(T)-phi(A) (<CR> for none): visc-a1_fa

Current dataset contains 22795 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 [visc-a1_fa.ins]: visc-a1_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.

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

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

As for the wavelength, the default value (as it is Cu K α) is correct. For a SAD experiment, the wavelength and the specified atom type are unimportant! Give a resolution cut-off of 2.3 Å.

```
Wavelength [1.54178]: 1.54178
```

```
Resolution cutoff in Angstroms [ 1.5 ]: 2.3
```

The instructions for SHELXD are displayed as given in the file. Can you find the resolution cutoff? Where is the wavelength?

To be later able to run SHELXE, 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: visc-a1.hkl. The R(free) flags are only useful if you intend to refine later with SHELXL.

You can exit the program, and proceed with SHELXD.

Finding the marker atom substructure with SHELXD

Just run from the command line:

```
shelxd visc-a1_fa
```

Or, to use the super-fast new beta-version of SHELXD for multiple processors:

```
shelxd_mp visc-a1_fa
```

Output files from SHELXD are:

visc-a1_fa.res marker atom solution with the best CFOM; this file is updated while SHELXC runs

visc-a1_fa.lst log file

Did it work?

The **best CFOM value** gives an initial indicator on solutions. It should be at least over 35 a good S-SAD solution. For a good solution, CC_{weak}, CC_{all} and CFOM should be high!

Have a look to the file visc-a1_fa.res! What are the different columns?

SHELXD always tries to find more marker atoms than requested in the FIND instruction. This is why you can get more sites in the output. A sharp drop in occupancy between site 6 and 7 can indicate a good solution. However, it can be difficult to discern correct solutions at this stage. You should process the best solution (in terms of CFOM) to SHELXE.

How could you even get more high-occupancy marker atom positions than you expected in S-SAD?

Density modification and auto-tracing with SHELXE

SHELXE carries out density modification and auto-tracing of the protein backbone. With these good data, all other parameters are sufficient at default values. Even the solvent content defaults to 45%, but here we expect 50%.

(The volume of the rectangular unit cell is approximately 66 Å x 66 Å x 47 Å = 204732 Å³. 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 Å³ of space. We expect two chains à 46 amino acids in the asymmetric unit, 92 in total. They would hence occupy 92 x 140 Å³ x 8 = 103040 Å³. 103040 Å³ divided by 204732 Å³ is approximately 0.50, which can then be the protein content. 1 - 0.5 is 0.5 which is the solvent content.)

Please use these two commands in separate windows:

```
shelxe visc-a1 visc-a1_fa -a -h -s0.4
shelxe visc-a1 visc-a1_fa -a -h -s0.4 -i
```

This will read in

```
visc-a1.hkl    native data
visc-a1_fa.hkl  anomalous differences,  $\alpha$ 
visc-a1_fa.res  marker atom substructure
```

At this stage, we cannot know if the substructure or its mirror image is the correct solution. The `-i` option ensures that both enantiomorphs are tested.

NOTE: SHELXE input follows the general form

```
shelxe XX YY <options>
```

XX is the native data set

YY is the experimental data file prepared by SHELXC or XPREP.

The output files are:

```
visc-a1.lst    Log file
visc-a1.pha    Anomalous density map
visc-a1.pdb    Backbone trace with 92 traced residues (out of 92!)
visc-a1.hat    Improved substructure coordinates (hat stands for heavy atom solution)
visc-a1.phs    hkl file with phases to display electron density
```

Running SHELXE will take some time. Please take care to study the output! What information is in there?

Did it work?

If the auto-tracing actually worked can be judged by the '**CC for partial structure against native data**' in SHELXE. In several rounds of auto-tracing, it should vary at a low value, increase suddenly and then vary at a higher value, which is higher than 25% for a successful trace at resolutions up to 2.2Å. Here, it should rise over 25% in the second cycle of auto-tracing.

How do we discern the correct hand after running SHELXE? Look at the values for the CC for the partial structure against native data. If you did not autotrace the structure, the correct hand is the one with the higher contrast which is printed to the screen at each cycle of density modification. This contrast can also be given graphically by HKL2MAP, as well as the less reliable map correlation coefficient. If in doubt: The correct map should look less ragged after some cycles of density modification. Clear side chains protruding from a traced backbone are an unambiguous sign of right hand!

You should use **COOT** to look at your results: Open the PDB file in COOT and then open the pha file. Therefore you have to use the option 'Open MTZ, mmCIF, fcf or phs...'. You will be asked if the cell from the PDB file should be used for the map, and you confirm this. To see if the anomalous density confirms the positions from the hat (and PDB) file, heighten the sigma value for the map. You can also look at the autotraced structure in the PDB file or at the normal electron density in the phs file. Which elements/which residues may account for the marker positions? This density is a good, mostly model-bias free starting point for building a model as complete as possible before your first REFMAC refinement. It can be enhanced a little by having a free lunch in SHELXE.

Selected SHELXE options

- h** Use this option if the marker atoms are part of the native structure and data. If their number is known, append it to h, e.g. `-h12`.
- a** Do auto-tracing of the backbone.
- q** Explicitly search for helices.
- i** Invert the enantiomorph of the substructure!
- s** Solvent content, set to 45% by default.
- e** free lunch, e.g. `-e1.2`
- z** optimize marker atom substructure before starting density modification

MAD

Molecule:	Concanavalin A
Number of amino acids	237
Best resolution	1.68 Å
Space group	I222
Cell	a=61.90Å b=87.06Å c=88.92Å $\alpha=90^\circ$ $\beta=90^\circ$ $\gamma=90^\circ$
Strategy	Three-wavelengths MAD data set

Data measured by Marianna Biadene and Ina Dix at DESY.

See also Hardman K.D., Ainsworth C.F. *Journal: Biochemistry* 11 (1972) 4910-4919

Data preparation with SHELXC

The input text file should be given as:

```
CELL 61.90 87.06 88.92 90 90 90
SPAG I222
FIND 1
NTRY 100
PEAK conca_peak.sca
INFL conca_infl.sca
HREM conca_hrem.sca
```

Cell parameters
Space group
Marker atom number
Number of tries (Good data, 100 sufficient)
Filenames of data sets; **NAT filename.sca** could be a native data set with better resolution; LREM is also possible.

If you are using XDS format, just give the files from different integrations different names – SHELXC does not rely on the filename extension to determine the format!

```
shelxc conca_mn < yourinputfile.inp | tee shelxc.log
```

(The command 'tee' prints the output both to the terminal and to the file shelxc.log, so that the program output can be kept for further reference. '| tee shelxc.log' can be omitted if you run SHELXC on a Windows machine.)

The last table in the log file gives correlation coefficients (%) between signed anomalous differences. As the data sets have to have consistent relative Bijvoet differences, this is an indicator of real anomalous signal or noise:

Correlation coefficients (%) between signed anomalous differences												
Resl.	Inf	- 8.0	- 6.0	- 5.0	- 4.0	- 3.5	- 3.0	- 2.8	- 2.6	- 2.4	- 2.2	- 1.96
HREM/PEAK	72.7	74.8	68.8	54.4	59.6	60.2	55.5	45.1	47.4	36.2	34.5	
HREM/INFL	72.4	74.3	67.1	48.9	60.0	55.8	49.7	42.8	38.6	29.6	34.5	
PEAK/INFL	94.5	91.3	90.6	84.8	84.6	79.0	70.7	63.5	53.9	50.8		

This correlation is so high that it is not necessary to truncate the data. If the data are of lower quality, they should be truncated where best correlation coefficient drops below 25%. To do so you have to change the **SHEL** line in the file name_fa.ins correspondingly, which was written by SHELXC.

Alternative data preparation with XPREP

Read in one of the data files (usually the one with the best resolution). If you are using XDS format, just give the files from different integrations different names – XPREP does not rely on the filename extension to determine the format! Proceed as before by choosing the correct lattice and space group. Then you choose from the main menu '[D] Read, modify or merge DATASETS', and that is the point at which you should read in the other data sets you want to use. For this, you should use the option '[R] READ in another dataset'. Since the cells are very similar, in this space groups the indexing is unambiguous – so just use the default options. The other data sets appear in the top list of the datasets menu; the current data set is marked with '<- current dataset'.

After you have read in the data sets, you can evaluate the anomalous signal in the data sets as before. If you want to change the current data set (to which the statistics you display will refer), use the option '[C] change current data set'.

When you are finished with the evaluation of the data, choose the option '[A] MAD, SAD, SIR or SIRAS' and then '[M] MAD (Multiple-wavelength Anomalous Dispersion)'.

Index # Data Filename or Source of Data

```

1 26975 conca_hrem.sca
2 26240 conca_peak.sca
3 20492 conca_infl.sca <- current dataset

```

```

[M] Sort-MERGE current data (no scaling)      [C] Change CURRENT dataset
[L] LEAST-SQUARES scale and merge datasets    [W] WRITE dataset to file
[I] INCLUDE Rfree flags from another file     [R] READ in another dataset
[S] Display intensity STATISTICS              [D] DELETE stored dataset
[F] FACE-indexed absorption corrections       [P] PSI-scan absorption corr.
[T] Copy file, TRANSFORM hkl and cosines     [A] MAD, SAD, SIR or SIRAS
[H] Apply HIGH/low resolution cutoffs        [N] NORMALIZE/scale sigmas
[G] Generate simulated powder diagrams       [U] Anisotropic scaling
[Y] Self or cross correlation coefficients     [J] Change dataset label
[Z] Expand data to triclinic                 [E] EXIT to main menu
[X] Parsons Q values and Flack x parameter   [Q] QUIT program

```

Select option [S]: **A**

```

[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]: **m**

The target number of reflections in local scaling sphere can be the default 100 for almost all cases. You should then give the first data set to process. It is good practice to start with the peak data, which has number 2 (see list above):

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

Number of remote or native dataset to which rest will be scaled [1]: **2**

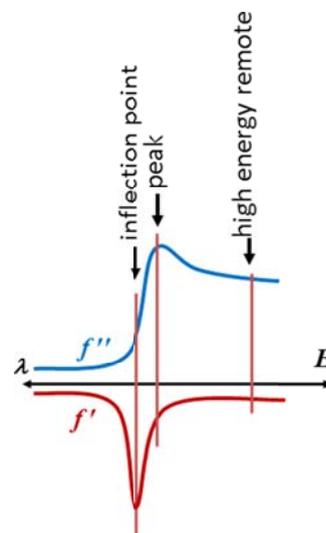
The program now asks for f' and f'' values. They actually are:

Data set	λ	f'	f''	data resolution
Inflection point	1.89360 Å	-8.0	2.5	2.30 Å
Peak	1.89240 Å	-5.1	3.9	1.96 Å
High energy remote	1.64219 Å	-1.0	3.1	1.68 Å

But we will behave as if we do not know. Just think about f' and f'' against resolution. The f' values are for almost all cases negative, and f' behaves like the first derivate of f'' . At inflection point, f' will be at its lowest value. As f' and f'' values can be put on an arbitrary scale, we decide that this minimum should be -8. Hence, at peak wavelength, it may be around -4 and at high remote -2 or -1.

The f'' has a maximum at the peak wavelength, and we assign a value of 7 to it. It will be lower at inflection point, around 4 and even lower at high energy remote, perhaps 3. We then have estimated:

Data set	f'	f''
Inflection point	-8	4
Peak	-4	7
High energy remote	-2	3



We will now use these estimated values for our SHELXD file preparation. It is of utmost importance to use the correct data sets! If inflection point and peak wavelength measurements are accidentally swapped, SHELXD may be able to find the marker atom substructure, but because of the wrong α values, SHELXE will not be able to get proper phases for the map, and density modification may not work!

Note: If you have suspicion that the data sets have been confused, you should use SHELXC (as described in the SHELXC tutorial) and the additional keyword SMAD. The data sets should be specified as usual for MAD. SHELXC will then prepare SAD phasing files from the MAD data and hence not use f' . The α values will be approximate but the data set confusion will be less important. You can also have a look in the header of your images and re-integrate your data. Be careful – MAD data set confusion is a common error.

Enter f' and f'' for this wavelength: -4 7

```
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 - 3.2 - 3.0 - 2.8 - 2.6 - 2.4 - 2.2 - 2.0 A
    1.32  1.53  1.39  1.34  1.47  1.46  1.49  1.54  1.51  1.36  1.21  0.98

    57.6  Neighbors used on average for F+/F- local scaling
Ranom =    0.0586  before and    0.0567  after local scaling
```

We then specify the next data set, inflection, and give again our estimated f' and f'' values:

Number of next MAD dataset (<CR> if none): 3

Enter f' and f'' for this wavelength: -8 4

```
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 - 3.7 - 3.5 - 3.3 - 3.1 - 2.9 - 2.7 - 2.5 - 2.3 A
    1.23  1.35  1.23  1.21  1.41  1.37  1.29  1.24  1.31  1.31  1.28  1.20

    58.9  Neighbors used on average for F+/F- local scaling
Ranom =    0.0446  before and    0.0426  after local scaling

    58.9  Neighbors used on average for local scaling to native/remote dataset
Ranom =    0.0195  before and    0.0195  after local scalin
```

```
Anomalous correlation coefficients (%) against previous datasets
Inf - 8.0 - 6.0 - 5.0 - 4.0 - 3.7 - 3.5 - 3.3 - 3.1 - 2.9 - 2.7 - 2.5 - 2.3 A
    95.1  91.5  89.9  82.9  81.6  84.2  80.9  75.1  67.8  60.5  53.4  36.6
```

The **anomalous correlation coefficient** is a much more reliable quality criterion of the anomalous signal than d''/σ because it does not require the $\sigma(I)$ values to be on an absolute scale! Any data for which it is higher than 25% should be used. Here, that would be all data. Running SHELXD with a high resolution means the resulting positions will be more accurate, but also, the substructure search will take longer.

Read in the high energy remote data as well, and give f' as -2 and f'' as 3. You can now refine f' and f'' values, which is usually not necessary. It is sufficient to refine each wavelength twice (e.g. in the order 123 123).

Number of next MAD dataset (<CR> if none):

Set	f'	Rf'	f''	Rf''	after MAD fit
1	-4.000	0.0342	7.000	0.0178	conca_peak
2	-8.000	0.0147	4.000	0.0311	conca_infl
3	-1.000	0.0196	3.000	0.0000	conca_hrem

It is not possible to refine all f' and f'' simultaneously, but with very precise data it may be worth refining individual values against the rest

Enter N to refine f' and f'' of set N or <CR> for no (further) refinement:

The weighting factor is for almost all cases 1.0; B is almost always good at 20.

The program then requires a filename, which should end with _fa.hkl to comply with the SHELXC standard (for example, vsc-a1_fa.hkl). Choose to write an .ins file for SHELXD and give it the same first part of a file. Specify the expected marker atom type (here Mn), and the expected number (1).

```
Weighting factor for dispersive contributions [1.0]: 1.0
Enter effective B-value (e.g. 20) to normalize delta-F or Fa values,
<CR> for no renormalization: 20
Filename to write FA and phi(T)-phi(A) (<CR> for none): conca-mn_fa
Current dataset contains 14152 MAD Fa-values
Write .ins file for SHELXD/XD (Y or N)? [Y]: y
Filename [conca-mn_fa.ins]:
Element type for heavy atoms [Se]: Mn
Number of unique heavy atoms [12]: 1
```

The wavelength you specify is not used by SHELXD or SHELXE. If you have native data against which you want to refine later, specify their wavelength. If that is not the case, you can give the peak data set's wavelength as an approximate average.

Give a resolution cut-off of 2.3 Å as a compromise between running time and accuracy.

```
Wavelength [1.54178]: 1.89
Resolution cutoff in Angstroms [ 2.4 ]: 2.3
```

You should write a native hkl if you want to run SHELXE later.
After this, you can exit the program, and proceed with SHELXD.

SHELXD

```
shelxd mp conca_mn_fa
```

In the output, you can see that the best CFOM (which is $CC_{All} + CC_{Weak}$) quickly rises above 60%, above one can be rather confident that the MAD solution in SHELXD is correct, so continue to SHELXE.

> **While SHELXC runs, try to estimate the solvent content.** (Hint: In I222, there are 8 ASUs per unit cell.)

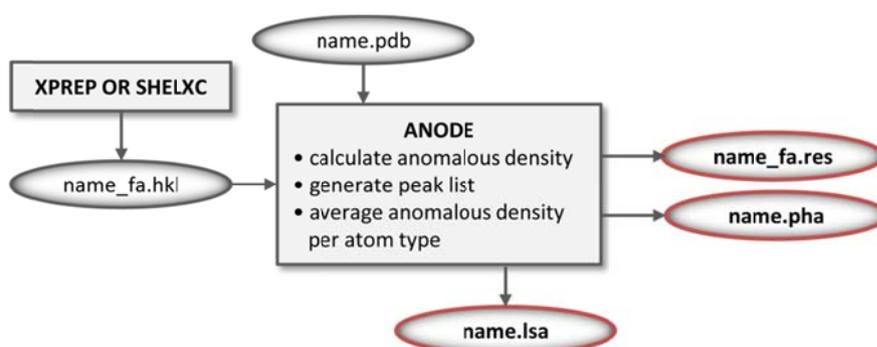
SHELXE

```
shelxe conca_mn conca_mn_fa -a
shelxe conca_mn conca_mn_fa -a -i
```

Here, no other options are needed. (The solvent content is 36% which is close enough to the default.) Again, evaluate your results in COOT. Can you optimize the last SHELXE steps with other options? (A list of options is given on top of the output when SHELXE is run!)

Extra tasks

- The folder co-conca contains Cu in-house data for SAD phasing of Concanavalin A with Cobalt. Try to solve the structure!
- Validation with anomalous density in ANODE: To validate the anomalous scatterer positions in your structure, and to find anomalously scattering ions in the hydration shell of your protein, you can use the tool ANODE. It will calculate the anomalous density for a PDB file.



Try it! In the folder visc-a1 is a final model of visc-a1 from the PDB and the file visc-a1_fa.hkl you generated. Now rename 3C8P.pdb to visc-a1.pdb and give:

```
anode visc-a1
```

You get an output on screen and can read in vsc-a1.pha to COOT, just as the pha file from SHELXE, but now it's based on the phases from the final refined structure. Look at the sulfates.

- Try having a free lunch with Concanavalin A, expanding the data to 1.2 Å. Compare the maps in COOT. Which one would you take to start building a model, and why?
- Try solving both structures with HKL2MAP and make use of the graphical output of HKL2MAP!
- What happens if you give a better resolution by the SHEL command in SHELXD and also heighten the number of NTRY, for example NTRY 1000? Do the positions of the anomalous scatterers improve? Do solutions occur more frequently? (Hint: This is most comfortably done with the shelxd_mp, the multi-processor version of SHELXD, as 1000 tries require some time.) Does this lead to better phases?