Solving structures with SHELX

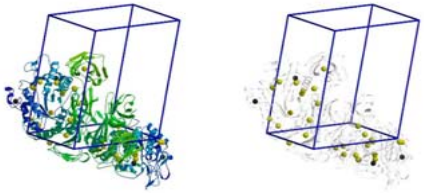
Roland Pfoh

rpfoh@uhnres.utoronto.ca

Saskatoon Synchrotron Summer School V
June 15, 2010

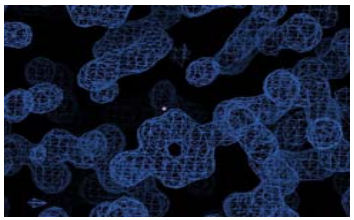
Programs involved

- SHELXD (substructure determination)



Tim Grüne, Göttingen

- SHELXE (density modification, main chain model building [integrated in beta test version], calculation of electron density maps)



Experimental density from Sulfur-SAD, calculated with 1.3 Å native data, displayed in Coot.

Aim of SHELXD/E

The intention is to get reasonable phases fast, in order to get an indication whether the structure is solvable before leaving the synchrotron. Priority is on speed, robustness and ease of use.

The programs were also designed for high throughput structure determination pipelines.

Often the resulting map is good enough for wARP to trace.

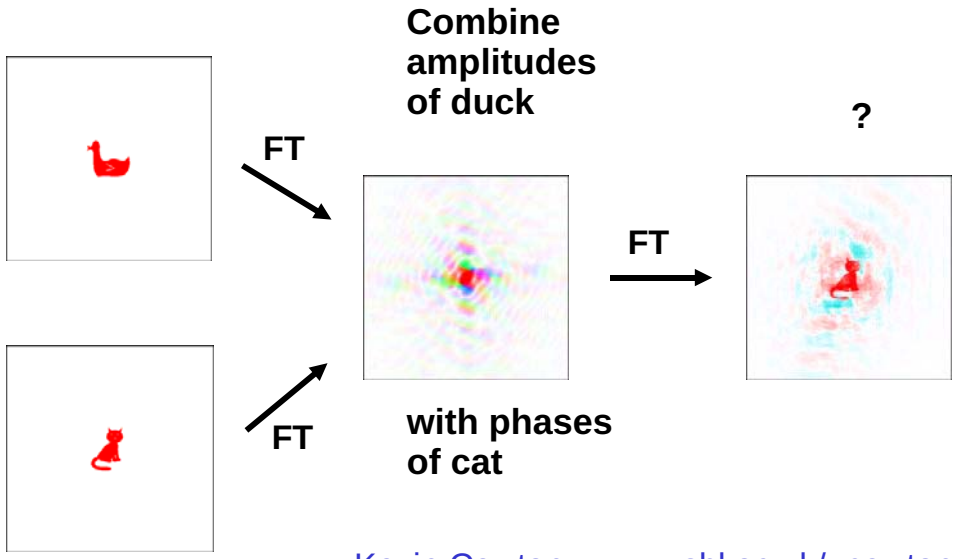
Some theory behind SHELXD

- In the anomalous diffraction experiment only the amplitudes of F(+h+k+l) and F(-h-k-l) (the Friedel pairs) can be measured.
- What we would like to know is F_A, which is the contribution of the heavy atom substructure.
- Depending on which experiment was done (MAD, SAD, ...), F_A can be estimated from F(+h+k+l) and F(-h-k-l) with different mathematical procedures.
- With the knowledge of F_A, we can artificially create a small molecule dataset of the heavy atom substructure, which is solvable by so called “direct methods”.
- With the knowledge of the heavy atom substructure the electron density of the complete macromolecule can be estimated. In the successful case this leads to an interpretable electron density map.

Some theory behind SHELXE

- These phases are then improved by **density modification**. If one does an inverse Fourier transform of the unmodified density one would get back those initial phases.
- Density modification implies a chemically sensible modification of the density before doing inverse FFT. The simplest idea (but quite useful) is truncating negative density to zero.
- In SHELXE the **sphere of influence algorithm** is used to decide whether a pixel in the electron density map belongs to protein or solvent. Here, the variance V of the density on a spherical surface of radius 2.42 A is calculated for each pixel in the map. The pixels with the highest V are most likely to correspond to real protein atomic positions.
- Initial estimates of the protein phases can be obtained by adding phase shifts to the heavy atom phases.

Amplitudes (F) and phases (ϕ)



Kevin Cowtan www.ysbl.ac.uk/~cowtan/

Input files for SHELXD and SHELXE

- Three input files are required:

`name-ha.ins` is the instruction file for SHELXD.

`name-ha.hkl` is in principle a small molecule dataset from the heavy atom substructure (contains F_A values) obtained from the anomalous experiment.

`name.hkl` contains native data, this can be either from the same crystal that gave the anomalous data or from a different crystal that diffracted to a higher resolution. For the latter case the two crystals should be fairly isomorphous, both alternatives can be tried. The first option has the advantage of perfect isomorphism, the second one might result in a much easier interpretable map.

- The first two files are read by SHELXD, the last two by SHELXE. SHELXE reads a file `name-ha.res` written by SHELXD, which contains the coordinates of the heavy atom substructure.

Preparation of input files

Starting from integrated and scaled diffraction data, there are different possibilities to generate input files:

- **XPREP** is a convenient program to analyze diffraction data and prepare files. It is the only program mentioned in this presentation that is not freely available, but the Bruker company usually provides a free version with an expiry date. It is also very useful to find the correct spacegroup.
- **SHELXC** is the standard program for generating the required files.
- **HKL2MAP** is an intuitive GUI that runs SHELXC, SHELXD and SHELXE and additionally generates useful graphical output.

How to run SHELXD/E

- All SHELX programs come as independent executable files and can be run directly from a terminal.
- SHELXD

```
SHELXD name-ha
```

- SHELXE (use two different terminals)

```
SHELXE name name-ha -h -m10 -s0.55
```

```
SHELXE name name-ha -h -m10 -s0.55 -i
```

-h: use if heavy atoms are present in native data.

-m: number of density modification cycles.

-s: solvent content.

-i: forces inversion of the heavy atoms in 2nd run (absolutely necessary after each new SHELXD run).

Critical parameters in SHELXD

- Correct spacegroup (no need to distuingish between enantiomeric ones, eg. $P4_12_12$ and $P4_32_12$).
- Resolution cutoff (Only data with a significant difference of the Friedel pairs should be used). Sometimes it is better to cut the resolution back.

```
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 - 2.5 - 2.1 - 1.9 - 1.7 - 1.5 - 1.3 A

   5.66  7.70  5.72  4.35  2.99  3.68  3.53  2.69  2.26  1.74  1.34  1.17
   6.45  8.49  6.43  5.04  3.26  3.74  3.43  2.42  2.06  1.67  1.35  1.22
```

Output from XPREP for Sulfur-SAD data.
The ratios should be higher than 1.3.
Possible resolution cutoffs would be 1.7 or 1.5 Å.
(You can consider the second line as R value for merged intensities, with the difference that you wish a high value for it.)

More Critical parameters in SHELXD

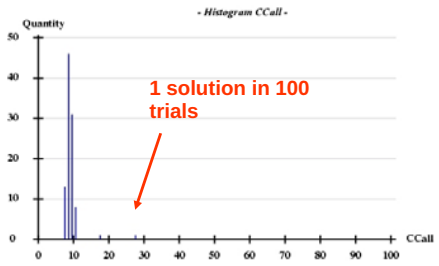
- Number of heavy atoms to search. One can expect a sharp drop in the occupancies of the found sites for Se-MAD and a gradual drop in the case of halide soaks.

In difficult cases it makes sense to vary those parameters (by editing the instruction file) systematically with a high number of tries (1000 or more) for each method. Auto tracing in the end could be used as validation.

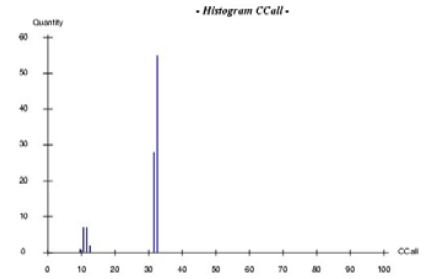
Values to look at in SHELXD

- A high value for the correlation coefficient {CC and CC(weak)} indicates a correct solution (eg. 30 and 15).
- The CC-values of correct solutions are usually well separated from the ones from wrong solutions.
- The best way to check is to run SHELXE with the best solution and look at the electron density map.

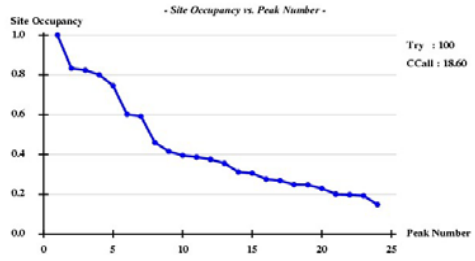
SHELXD histograms and occupancies for Elastase



S-SAD

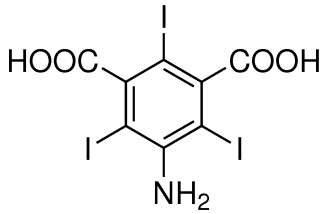


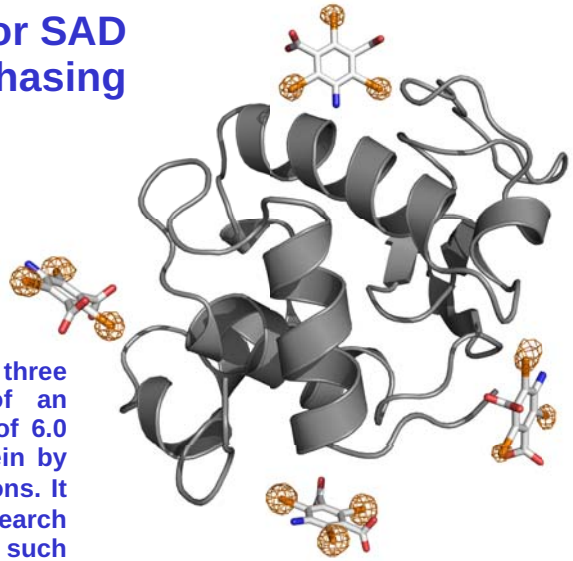
I-SIRAS



George M. Sheldrick, Göttingen

A magic triangle for SAD phasing





The above molecule contains three iodine atoms in the form of an equilateral triangle with a side of 6.0 Å. It can attach itself to a protein by hydrogen bonds and π-interactions. It is intended to adapt the peak-search in SHELXD to search for such triangles to give enhanced SAD phasing despite partial occupancies.

Tobias Beck, Göttingen

Remark: The magic triangle is also available with bromine instead of iodine for MAD phasing.

Critical parameters for SHELXE

- Number of cycles (-m10 up to -m200).
- Solvent content (-s0.50, -s0.55, ...); optimal value can be calculated using the cell volume, the number of proteins to expect in the unit cell (check Wyckoff positions in the International Tables) and the protein size [the volume of a non-hydrogen atom is 18 Å^3].
- Native dataset; the resolution should be high and the crystal should be as isomorphous (similar cell constants) as possible with the crystal used for the anomalous experiment (e.g. same crystal).
- Free lunch algorithm (if the native dataset has a resolution <1.7 Å try -e1.0, otherwise try to improve it by 0.2 Å).
- Autotracing; can help in difficult cases. Try either free lunch or autotracing.

Values to look at in SHELXE

- Contrast and connectivity should be high and well separated from the wrong enantiomorph. Only one of the two SHELXE jobs can result in a correct map. In difficult cases the difference between the two enantiomeres can be smaller. Look at both maps in this case.
- Pseudo-free CC (computed at the end of each run) should be close to 0.7 or higher. In difficult cases this number might be smaller, but bellow 0.6 the map will unlikely be interpretable.
- Check the electron density map:
  1. You should see solvent channels (areas with less or no density).
  2. look for alpha helices (look depends on resolution).
  3. Sometimes the heavy atom position from SHELXD can help to identify a correct solution.
  4. For medium resolution the protein should have a more or less continuous density. Only for very high resolutions (<1.2) the correct map might look atomic instead of continous.

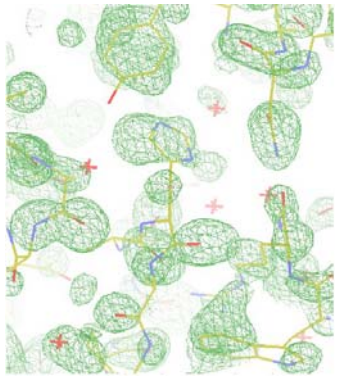
The free lunch algorithm

- This algorithm tries to extend the resolution of the measured data by inventing (!) intensities of unmeasured reflections and extrapolating their phases. It has been shown that the density maps improves in a lot of cases if the native data has been measured to at least 2 Å.

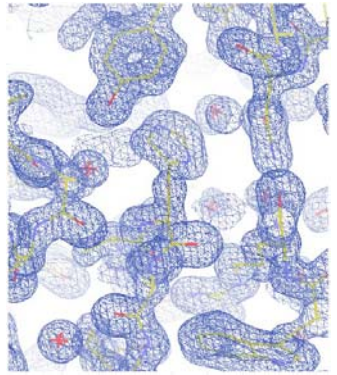
- Possible explanations of this unexpected result are:
  1. The algorithm corrects Fourier truncation errors.
  2. Phases are more important than amplitudes.
  3. Zero is a poor estimate of an unmeasured reflection.

- The name of this algorithm was chosen to express the feeling that sometimes accompanies an invitation for a free lunch: One can't bee sure if it's really for free!

Maps before and after a free lunch



Best experimental phases without free lunch (MAPCC 0.57)



After expansion to 1.0 Å with virual data (MAPCC 0.94)

The only phase information was a weak SIRAS signal to 3.5 Å from a mercury acetate derivative. The native data was relatively complete to 1.35 Å.

Isabel Usón, Barcelona

Beta test version for chain tracing in SHELXE

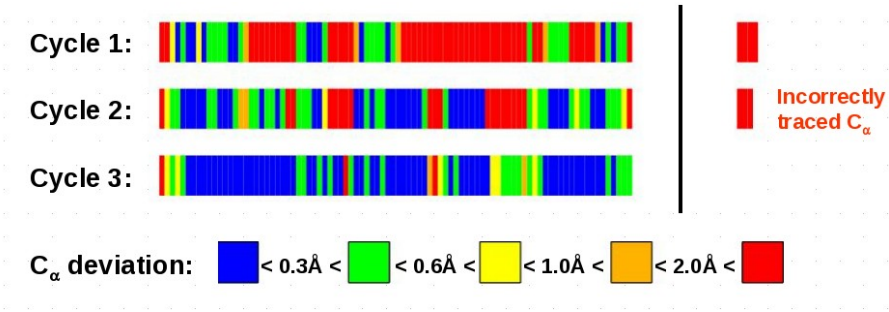
- A fast but crude autotracing function is currently tested in SHELXE.
- It is primarily designed to get a toehold with very poor starting phases and use other tracing programs afterwards.
- The optimal condition for testing would be the following: The structure is probably solved but the map quality is too bad for building a model.
- All that needs to be done for testing is to type -a3 (for 3 cycles of autotracing) after the standard SHELXE command line:

```
SHELXE name name-ha -h -m10 -s0.55 -a3
```

- In order to get the test version, please contact George Sheldrick and he will provide you with the download instruction. Visit the SHELX homepage for all sorts of information:

http://shelx.uni-ac.gwdg.de/index.html

Fibronection autotracing



George M. Sheldrick, Göttingen

This structure illustrates the ability of the autotracing to start from a noisy S-SAD map. Recycling the partial but rather accurate traces leads to better phases and an almost complete structure.

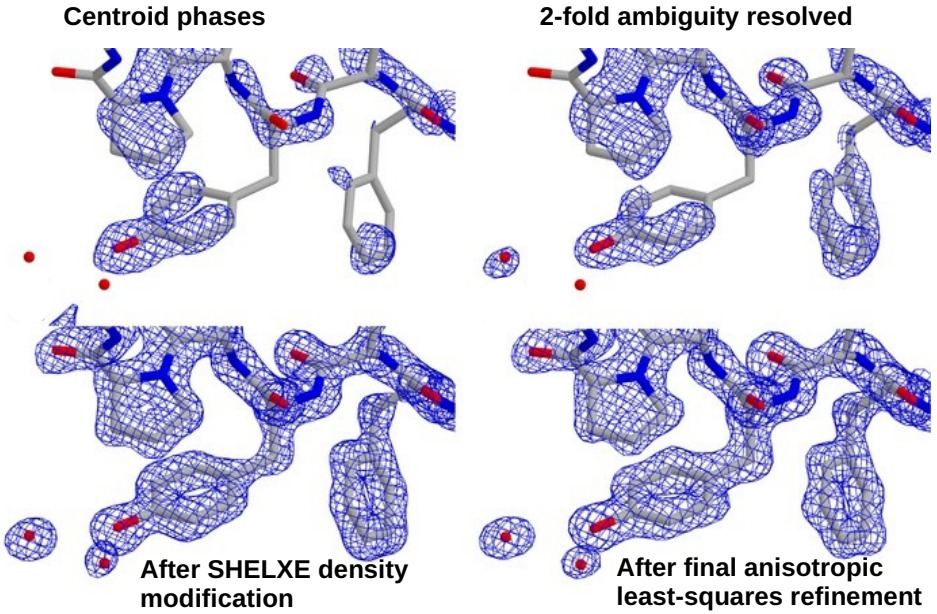
Fibronection map quality

	MPE[°]	mapCC
Standard S-SAD [−h −s0.35]:	53.4	0.63
S-SAD with FLA [−m200 −h −s0.5 −e1]:	42.9	0.70
S-SAD with autotracing [−h −s0.35 −a3]:	32.3	0.84
S-SAD, autotracing and FLA [−h −s0.35 −a3 −e1]:	31.6	0.86

George M. Sheldrick, Göttingen

An interpretable map can be obtained by either the free lunch algorithm or autotracing. A combination of both only leads to a tiny improvement. The free lunch algorithm has the advantage that it works also for compounds like DNA, RNA, PNA that are difficult to trace automatically.

Stages in the phasing of cubic insulin



George M. Sheldrick, Göttingen

Acknowledgments

- Prof. George M. Sheldrick (University of Göttingen, Germany).

- Prof. Emil Pai (University of Toronto) and his group.

UNIVERSITY OF TORONTO

HFSP