Macromolecular Phasing with SHELXC/D/E

Tim Grüne

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

tg@shelx.uni-ac.gwdg.de
Overview

**shelxc**: Data Preparation

**shelxd**: Substructure Solution

**shelxe**: Phase Improvement

- Sphere of Influence
- Phase Extension with Free Lunch
- Backbone building of Proteins
- Proposed Algorithm for Building Nucleic Acids
Purpose of *shelxc/d/e*

The *shelxc/d/e*-triad can be used for phasing by:

- anomalous dispersion (SAD, MAD)
- isomorphous replacement (SIR, MIR)
- combination of SAD and SIR (SIRAS)
- radiation induced phasing (RIP)

which covers about all phasing methods except for Molecular Replacement
Using \textit{shelxc/d/e}

\textbf{shelxc} merges and scales data sets and prints useful statistics (e.g. to judge resolution cut-off for \textit{shelxd})

\textbf{shelxd} solves the substructure, i.e. position of the anomalous/ heavy scatterers.

\textbf{shelxe} initially was designed to quickly check whether collected data are sufficient for solving the structure. Refined phases to start model building from were supposed to be prepared by “more sophisticated” programs (Sharp, DM, …). However, since \textbf{shelxe} turned out to produce phases suitable for model building, this “detour” is often not necessary.
Sample Run

The *shelx* programs are usually run from the command line. With very few command lines an electron density map can be prepared that can be fed into, e.g., ARP/wARP.

```bash
#> shelxc gere << eof
CELL 109.02  61.75    71.74   90.00   97.08   90.0
SPAG C2
PEAK  gere_peak.sca
INFL  gere_infl.sca
HREM  gere_hrem.sca
LREM  gere_lrem.sca
FIND  12
eof
#> shelxd gere_fa
#> shelxe gere gere_fa -s0.48 -m20 -b
#> shelxe gere gere_fa -s0.48 -m20 -b -i
```

*hkl2map* (T. R. Schneider, [http://webapps.embl-hamburg.de/hkl2map/](http://webapps.embl-hamburg.de/hkl2map/)) is a very helpful GUI that also visualises and summarises important output of the programs.

S. Rühl (SHELX group) is currently working on a new GUI which is going to be released soon.
Critical Parameters

**shelxd**
- Resolution cut-off: do not include data without signal
- **FIND** should be within 20% of the correct number of sites

**shelxe**
- Solvent Content: Try varying the theoretical content
Data from integration programs must be converted to .hkl, or .sca to be read in to shelxc. Depending on the integration software, different paths must be taken:

**HKL2000**  
.sca files can be read directly into shelxc.

**Mosflm/Scala**  
the program mtz2sca (SHELX) converts the output from Mosflm or Scala into a .sca-file. Please make sure truncate also keeps intensities. Alternatively: Mtz2 various

**XDS**  
two possibilities:

1. xds2sad (SHELX) and sadabs (Bruker/AXS), if available
2. Combat/ Scala (ccp4i GUI), then mtz2sca.
**shelxc: Output**

**shelxc** produces three files:

**name.hkl** the merged and scaled data of input. It is used by **shelxe** and later during refinement.

**name_fa.hkl** a file containing the anomalous differences, used by **shelxd** to localise the substructure atoms, and the phase angle $\alpha$ (definition below) used by **shelxe**.

**name_fa.ins** instruction file read by **shelxd**.

The terminal output contains information about data quality, especially the strength of the anomalous difference of the data:

```
Resl. Inf - 8.0 - 6.0 - 5.0 - 4.0 - 3.5 - 3.0 - 2.5 - 2.0 - 1.8 - 1.6 - 1.37
N(data)  275  350  420  955  933  1669  3212  7178  5453  8462  15819
<I/sig>  122.6 124.9 128.7 139.0 137.1 125.2  94.7  67.9  43.3  24.7  11.5
%Complete  94.2  99.2  99.3  99.6  99.6  99.8  99.8  99.9  99.9  99.9  93.0
<d"/sig>  2.32  2.80  2.01  1.80  1.58  1.52  1.39  1.25  1.01  0.91  0.82
```

Sample output of **shelxc** for a case of SAD.

**Rule of thumb:** choose resolution cut-off for **shelxd** where $<d"/sig> > 1.3$. 
Substructure Solution with \texttt{shelxd}— SAD

In the presence of an anomalous scatterer, the Bijvoet pair $|F(+hkl)|$ and $|F(-hkl)|$ differ.
Substructure Solution with shelxd— SAD

The differences between the Bijvoet pairs generate a Patterson map approximately as though only the substructure had been present in the crystal (with the same unit cell dimensions).

With each of its trials, shelxd generates a list of heavy atoms consistent with the Patterson function generated from the $F(±hkl)$ pairs and refines them with dual-space refinement.

The coordinates of the best solution are written to the file name_fa.res used by shelxe.
In this diagram, $F_{\text{mol}}(hkl)$ marks the contribution of the main molecule without the substructure (heavy atoms). $F_{\text{h.a.}}(hkl)$ is the non-anomalous contribution of the substructure atoms to the total structure factor, i.e. $F_{\text{tot}}(hkl)$ is the structure factor as it were without anomalous dispersion.
Phasing with \textit{shelxe}

SAD-phasing in \textit{shelxe} is based on a formula derived by Karle (1980) and Hendrickson, Smith, Sheriff (1985):

\[ |F(\pm hkl)|^2 = |F_{tot}|^2 + a|F_{h.a.}|^2 + b|F_{h.a.}||F_{tot}| \pm c|F_{h.a.}||F_{tot}| \sin \alpha \]

\[ a = \frac{f''^2 + f'^2}{f_0^2}, b = \frac{2f'}{f_0}, c = \frac{2f''}{f_0} \]

This leads to the approximation

\[ |F(\pm hkl)| = |F(\mp hkl)| \approx c|F_{h.a.}(hkl)| \sin \alpha \]

This is exploited by \textit{shelxe} in order to improve the phases.
Phasing with *shelxe*

In order to improve phases, *shelxe* requires starting phases.

In the case of SAD, *shelxc* sets the angle $\alpha$ according to:

\[ |F(+hkl)| > |F(-hkl)| \quad \text{set } \alpha = 90^\circ \]
\[ |F(+hkl)| < |F(-hkl)| \quad \text{set } \alpha = 270^\circ \]

These are the starting phases for *shelxe*. While these seem very rough estimates they are in most cases sufficient.

For SIR the approximation is similar, but leads to the angles $\alpha = 0^\circ$ and $\alpha = 180^\circ$.

With MAD or MIR, angle $\alpha$ can actually be calculated exactly within experimental error and the initial electron density map is usually of much higher quality than for SIR or SAD.
**shelxe: Density Modification with the Sphere of Influence**

**shelxe** calculates the **probability** of every point in the asymmetric unit whether it belongs to the (disordered) solvent region or the ordered region of the macromolecule. Ordered density is being enhanced whereas **solvent flipping** is applied to disordered density. The modified density is transformed back into reciprocal space by reverse Fourier transform. An improved density map produces improved phases.

To monitor this process, the command line switch -b lets **shelxe** produce a map of the anomalous density (.pha-file) readable by coot.

The border between solvent and disordered region is determined by the **Sphere of Influence**.
shelxe: The Sphere of Influence

How does shelxe decide, whether a pixel is part of the solvent or ordered region?

- Place a sphere of 2.42Å diameter at every point in the asymmetric unit.
- 2.42Å is the typical 1,3-distance in macromolecules.
- Calculate variance of electron density on 92 surface points.
  - large variance → protein region
  - small variance → solvent region
  - smooth transition in the “fuzzy” region

The idea behind the free lunch algorithm:

- extend the resolution of the reverse Fourier transform from map to structure factors
- roughly estimate structure factor amplitudes and use phases from reverse Fourier transform: a poor estimate is better than no estimate

This strategy has improved the density for several test structures and introduced new features to the map not visible before.

Unfortunately, data better than approximately 2Å are required for this method to work.
**shelxe: Backbone Tracing of Proteins**

One very special and powerful way of density modification is to actually build a model. The (yet unpublished) version of *shelxe* has the ability to carry out autotracing of the main chain of proteins.

1. Find potential α-helices and β-strands in the density and try to extend them at both ends.
2. Avoid Clashes with already traced atoms and heavy atoms.
3. Tidy up and splice the traces as required, applying any necessary symmetry operations.
4. Use the traced residues to estimate phases and combine these with the initial phase information using $\sigma_A$ weights, then restart density modification.

A refinement step assigns B-values per residue and to the heavy atoms. This scales down outliers/ false traces.
Example NovP: Quality of Main Chain

Complete (in-house) data to 1.9 Å merged with incomplete 1.35 Å synchrotron data, but SIRAS phases from Hg-soak very poor (3.5 Å). Originally solved with the free lunch algorithm, but autotracing can also get a foothold.

$C_{\alpha}$-deviation trace vs. final

Incorrectly traced $C_{\alpha}$
A Proposed Mechanism to trace Nucleic Acids

The latest development in the SHELX group is the ambition to *trace to Nucleic Acids*. Autotracing requires regular properties with rather tight distributions in order to create selection criteria.
Electron density maps of nucleic acids contain two prominent features

1. Bases and base stacking (flat shapes)
2. Strong scattering from Phosphate atoms ($Z_P = 15$ vs. $Z_C = 6$)

Wadley et al. showed that the RNA backbone is determined through two torsion angles based on $P$ and $C4'$, similar to the Ramachandran plot for proteins.

Because bases are better determined than the sugar ring, we concentrate on relations between the phosphate atoms and the base plane.
Angle between the Base Plane and the $5'P$-$3'P$ vector
Macromolecular Phasing with SHELXC/D/E

Tracing of Nucleic Acids
Asymmetric Distribution of $C1'$ to $5'P$ and $3'P$

Since the distance of $C1'$ to $5'P$ is expectedly shorter than $C1'$ to $3'P$, a direction can be assigned to a $P$-Base-$P$ triplet and hence to a chain.
Algorithm

1. Locate $P$-Atoms and Bases (see e.g. Hattne et al., Acta Cryst. (2008), D64 pp. 834-842)
2. consolidate bases by checking for Watson-Crick pairs
3. build $P \rightarrow C1'(\text{base}) \rightarrow P \rightarrow C1'(\text{base}) \rightarrow \ldots$ chain based on aforementioned asymmetries to fix $5' \rightarrow 3'$ directionality.
References

- [structbio.biologie.uni-konstanz.de/ccp4wiki](http://structbio.biologie.uni-konstanz.de/ccp4wiki) This Wiki contains a lot of shelx specific information, tips, tricks, and technical information.
Acknowledgment

Part of this work was supported through the EMBO fellowship ALTF-1071 (2003) and through the BIOXHIT Project.