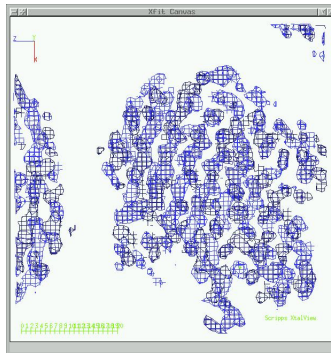# Tutorial for solving structures with HKL2MAP 0.1

Thomas R. Schneider

IFOM - The FIRC Institute of Molecular Oncology

www.ifom-firc.it

schneider@ifom-firc.it

October 26, 2003

```
jia_hrem.sca
jia_infl.sca
jia_lrem.sca
jia_peak.sca
```

$\rightarrow$

1 Intro

HKL2MAP is a graphical user-interface for macromolecular phasing. The current version connects several programs from the SHELX-suite to guide the user from analysing scaled diffraction data (SHELXC), via substructure solution (SHELXD) and phasing (SHELXE) to displaying an electron density map (Xfit). HKL2MAP can handle SAD, SIR, SIRAS or MAD data. This tutorial describes the determination of phases from a MAD experiment is described.

HKL2MAP was written by Thomas Pape and Thomas R. Schneider. All rights are reserved. The present version 0.1 is a beta-test version and can be obtained from `schneider@ifom-firc.it`. The authors do not take any responsibility for the correct function of HKL2MAP or any of the programs called by it. We do, however, appreciate reports on faulty, unexpected or destructive behaviour to `schneider@ifom-firc.it`.

Information about the SHELX-suite of programs can be found at `shelx.uni-ac.gwdg.de/SHELX`

Please cite the following references when you report the use of HKL2MAP:

HKL2MAP Pape T & Schneider TR (2003). personal communcation.

SHELXC Sheldrick GM (2003). Göttingen University.

SHELXD Schneider TR, Sheldrick GM (2002). Substructure solution with SHELXD, Acta Cryst. D58:1772-1779.

SHELXE Sheldrick GM (2002). Macromolecular phasing with SHELXE, Z. Kristallogr. 217:644-650.

XFIT McRee DE (1999). XtalView/Xfit - A versatile program for manipulating atomic coordinates and electron density, J. Struct. Biol. 125:156-165.

2 Diffraction Data

The data used for this tutorial originate from a four wavelengths MAD-dataset on a Se-Met derivative used to solve the structure of *Escherichia coli* thioesterase II [?] (TE-II). In addition to the MAD-data, a native data set to 1.9 Å resolution was collected on the S-Met form of the protein. The data were kindly provided by Zbyzsek Dauter and can be downloaded from `http://shelx.uni-ac.gwdg.de/~trs/mad/mad.html`.

The structure of TE-II was determined in space group $C222_1$ with unit cell constants of: $a = 96.0$ Å, $b = 120.0$ Å, $c = 166.1$ Å, $\alpha = \beta = \gamma = 90.0°$. The data were processed with DENZO and SCALEPACK. The scaled and merged data were stored in SCALEPACK .sca-format.

code	λ [Å]	d_{min} [Å]	filename
hrm	0.9747	2.5	`jia_hrem.sca`
pk	0.9787	2.5	`jia_peak.sca`
ip	0.9793	2.5	`jia_infl.sca`
lrm	0.9801	2.5	`jia_lrem.sca`
nat	n/a	1.9	`jia.hkl`

Table 1: Overview of data sets used in this tutorial. `jia.hkl` is in SHELX-hkl4-format, i.e. contains $h, k, l, I, \sigma(I)$ in asci-format.

The refined structure of TE-II (pdb-code 1C8U) contains 2 x 4 Se-atoms in a total of 2 x 285 amino acids. The solvent content is ca. 66%. More information about the structure and its determination can be found in: Li et al. (2000) Nat.Stru.Biol. 7:555.

3 MAD phasing of TE-II

3.1 Starting HKL2MAP and Entering the Project Name

HKL2MAP can be started from a UNIX-command line:

```
% hkl2map &
```

After closing the Welcome-Window, the Main-Window of HKL2MAP becomes visible (Figure **??**). At the top of the Main-Window, a Menu-Bar and an input line for the project name are located. For the different programs called by HKL2MAP, the interface provides Panels that can be openend and closed by clicking the respective boxes. The bottom part of the Main-Window consists of a Status-Window containing the output generated by the called programs and a line displaying brief context-sensitive help-messages.
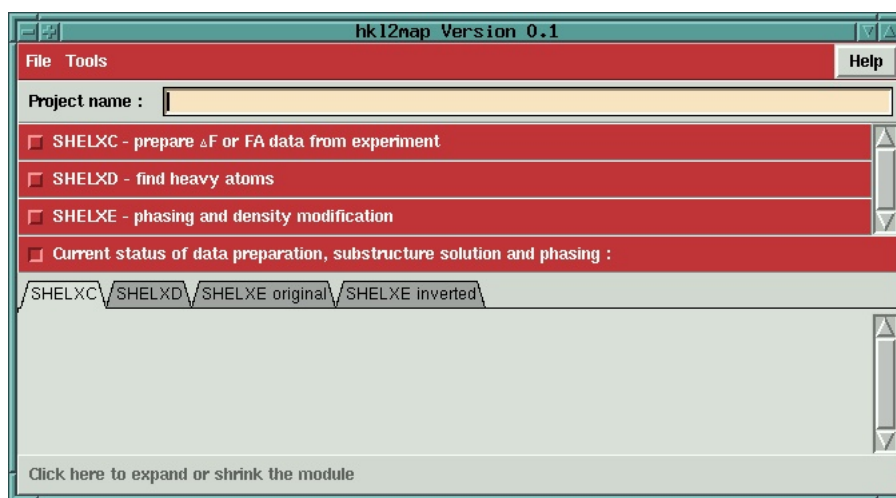

Figure 1: HKL2MAP Main-window after starting the program.

To get started, a project name needs to be defined. This name will be used as a prefix for all filenames used by HKL2MAP or the SHELX-programs called by it.

Once the project name has been entered, all input-fields of the interface concerning input- and output-filenames are loaded with default values derived from the project name, and the SHELXC-Panel is opened (Figure **??**).
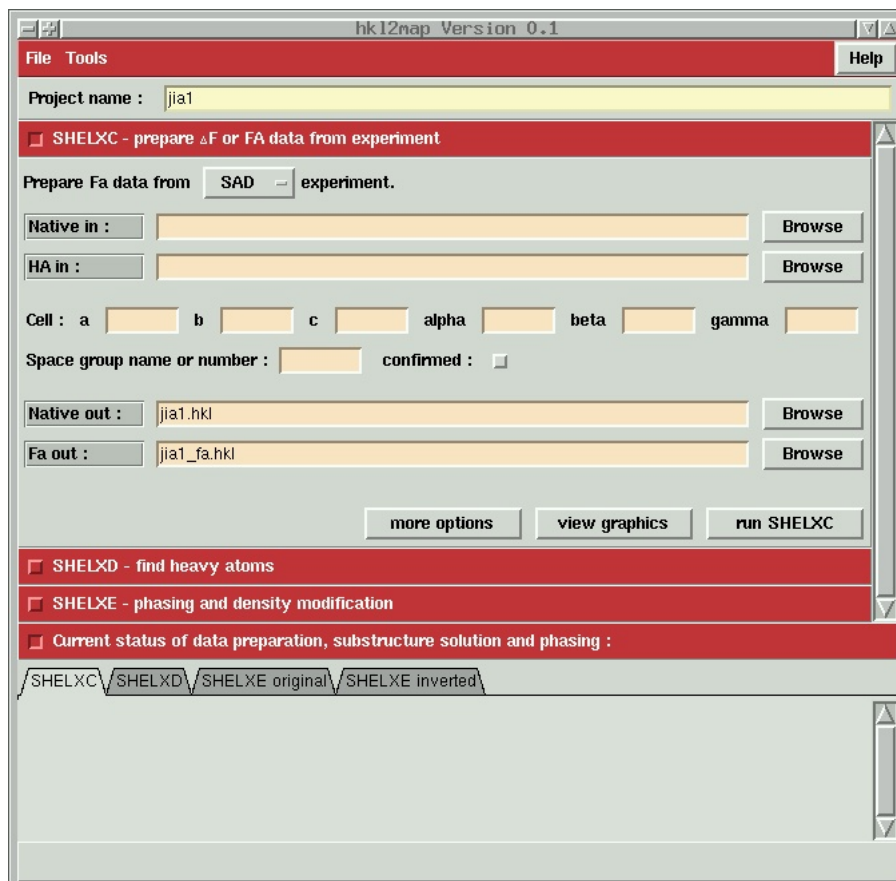
Figure 2: HKL2MAP Main-window after entering the project name.

3.2 Running SHELXC

To run SHELXC, first the type of experiment has to be chosen. After the choice has been made, the SHELXC-panel changes and presents input fields appropriate for the respective type of experiment. For the present example, MAD has been selected and, in response, the interface generates input fields for filenames for files containing the diffraction data for a native and four MAD data sets. The filenames can be typed in or files can be selected using a browser that is actived by clicking on the Browser-Button.
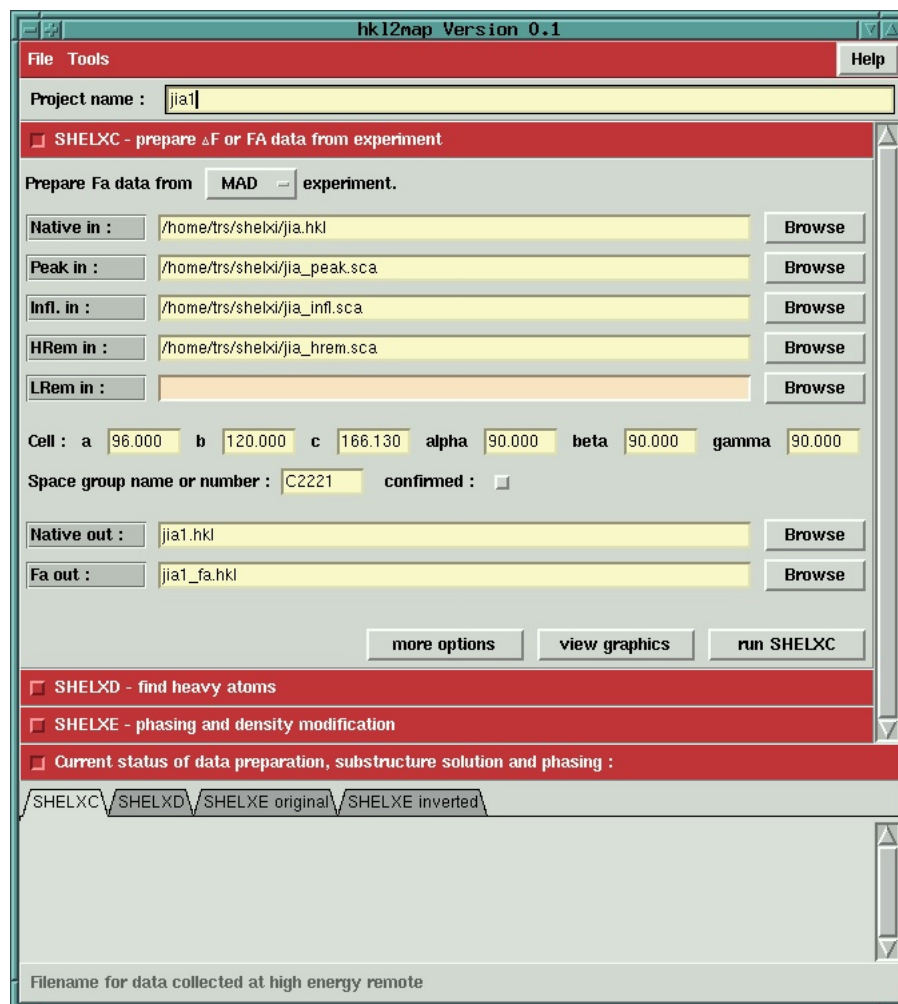

Figure 3: HKL2MAP Main-window after a MAD-experiment has been selected and names of files containing diffraction data have been entered.

As many 'accidents' happen due to misunderstanding between programs about space groups, HKL2MAP requires an explicit confirmation of unit cell constants and space group (Confirmed-Button) before SHELXC can actually be started from the interface.

Once SHELXC is started by clicking on the Run-SHELXC-Button, the interface is blocked for further input until the program has finished. The output of the SHELXC-run is directed to the Status-Panel at the bottom of the Main-Window.

After SHELXC has finished, a graphical version of the SHELXC-output can be activated by pressing the View-Graphics-Button. Clicking this button will open a graphics window allowing to display various plots. The type of plot to be displayed can be selected using the Display-Menu in the
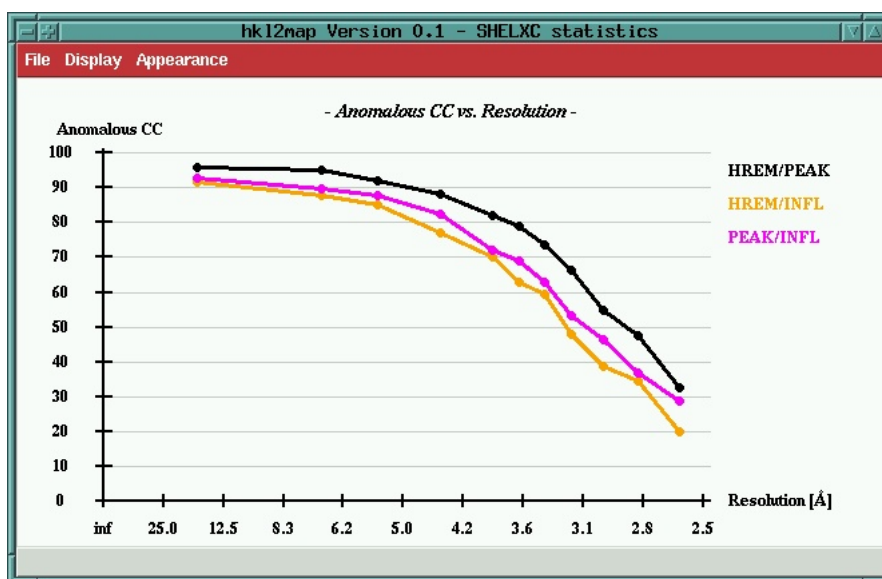
6

Graphics-Window.

Figure 4: The Graphics-Window for the results of SHELXC after choosing the plot of 'Anomalous CC vs. Resolution'.

The plots shown in the Graphics-Window can be written in postscript-format using the Make-Postscript-Option in the File-Menu of the Graphics-Window.

3.3 Running SHELXD

The name of the file containing substructure structures factors is automatically forwarded to the SHELXD part of the interface. Also, the interface tries to suggest a high resolution cut-off for the data used for finding the substructure.

The user has to supply the expected number of sites and the type of heavy atom in the crystal. In this case, 8 Se-sites are requested, and SHELXD is started by pressing the Run-SHELXD-Button.
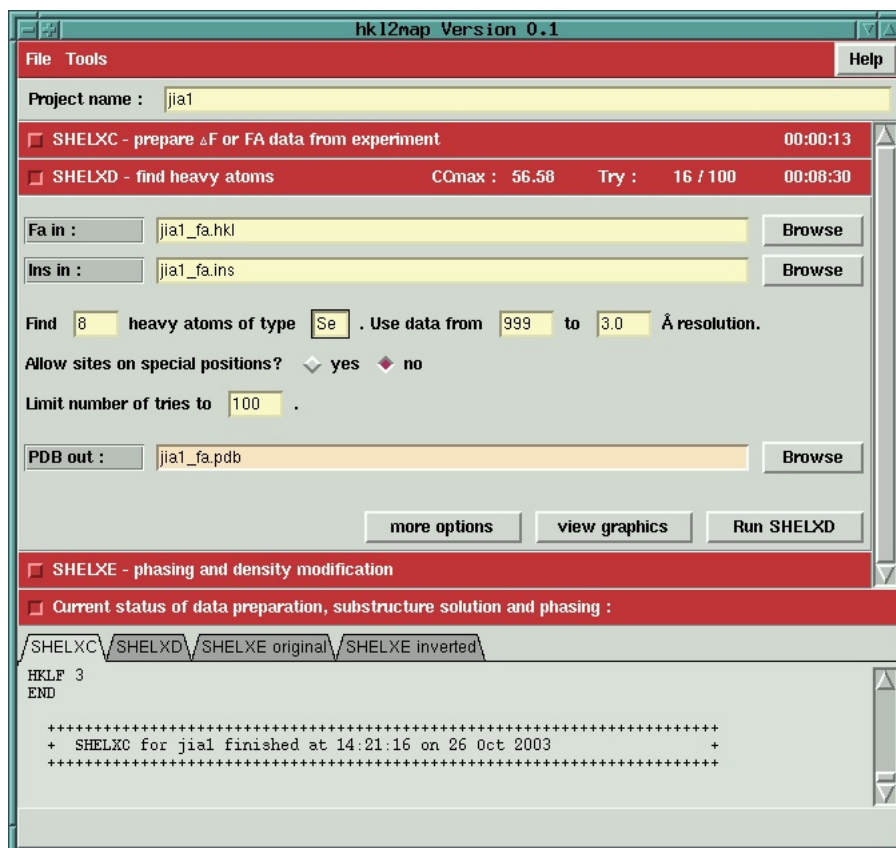

Figure 5: The Main-Window after entering the numbers necessary for starting a SHELXD-job.

The output from SHELXD is shown to the Status-Window at the Bottom of the Main-Window. Again, a graphical display of the results can be activated by pressing the View-Graphics-Button. As opposed to the SHELXC-related plots, the plots describing SHELXD are automatically updated to reflect the current status of the SHELXD-run and can be inspected at any time during the SHELXD-run.

One useful plot is the plot of peak-height vs. peak number for the current best solution (Figure **??**). In favourable cases, this plot shows a clear drop indicating the last significant site of the substructure.

A SHELXD-run can be interrupted at any time using the Finish-SHELXD-Button.
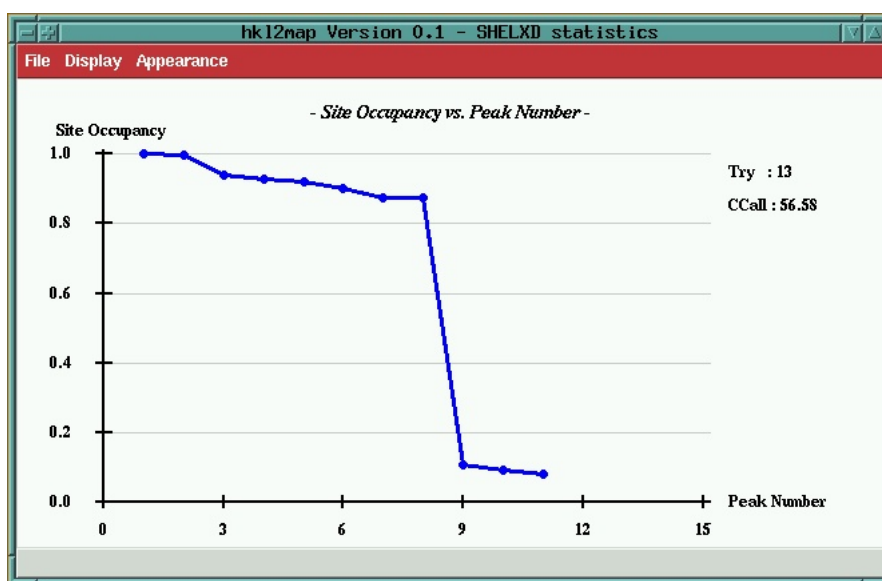
Figure 6: The Graphics-Window showing a plot of peak-heigt vs. peak-number. A clear drop in peak-height is seen between peak 8 and 9.

3.4 Running SHELXE

The names of the files containing the native data, the substructure structure factors and the best sites found by SHELXD are automatically filled into the respective fields of the SHELXE-panel of the interface.

Before calculating phases with SHELXE, the fractional solvent content (for convenience a widget for automatic calculation of the solvent content can be activated by pressing the Estimate-Solvent-Content-Button) has to be entered. Also the user has to supply the information whether or not the native crystals contained the substructure atoms or not.

The interface allows to specify which enantiomorph of the substruture should be used for phasing. As an alternative to testing the enaniomorphs one after the other, the respective jobs can be run simultaneouly by choosing try both enantiomorphs. Dual-CPU computers are particularly convenient to run the jobs for the two enantiomorphs in parallel.
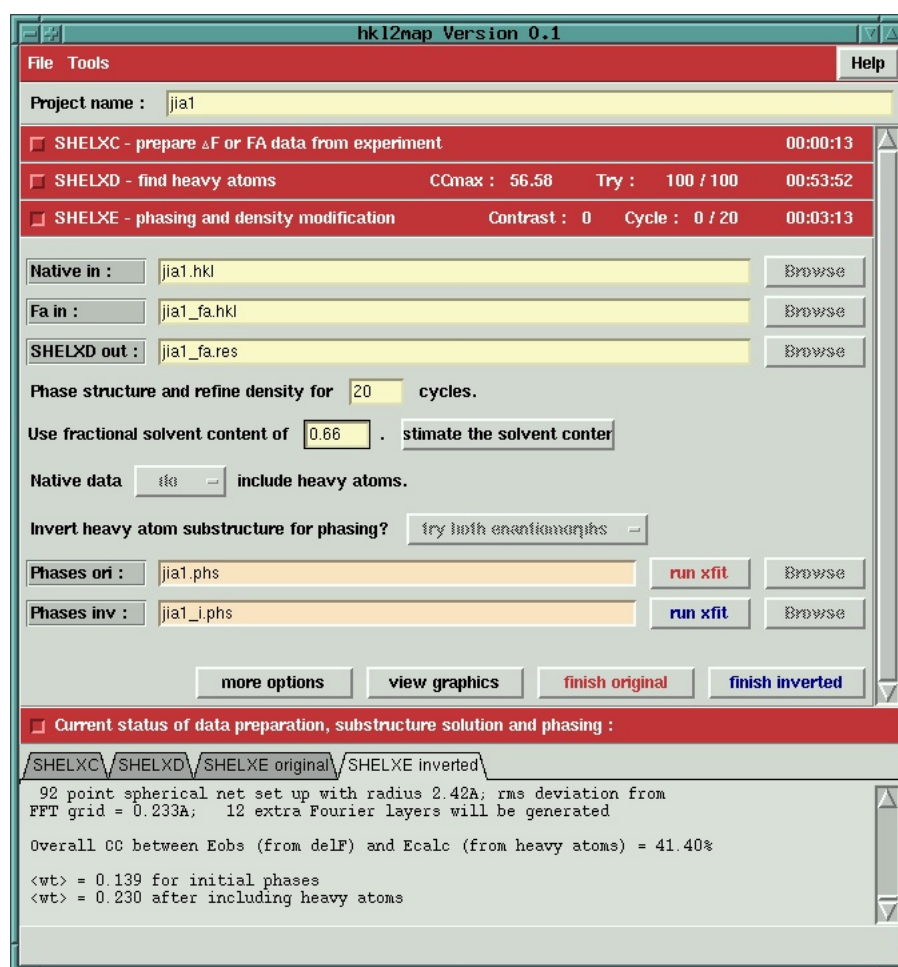

Figure 7: The Main-Window after starting SHELXE.

The progress of the phasing calculations can be followed in the Status-Window or, more conveniently, by activating the graphical view of the SHELXE-output by pressing the View-Graphics-Button. An important measure for phase quality is the 'Contrast' between the variance in the electron density in the protein region and in the solvent region for a given phase-set. Using the Graphics-Window activated via the View-Graphics-Button in the SHELXE-Panel, a continuously updated plot of the value of the contrast vs. cycle number can be displayed and the decision which

of the substructure-enantiomorphs results in the correct phases can be made while the phasing calculations are running.
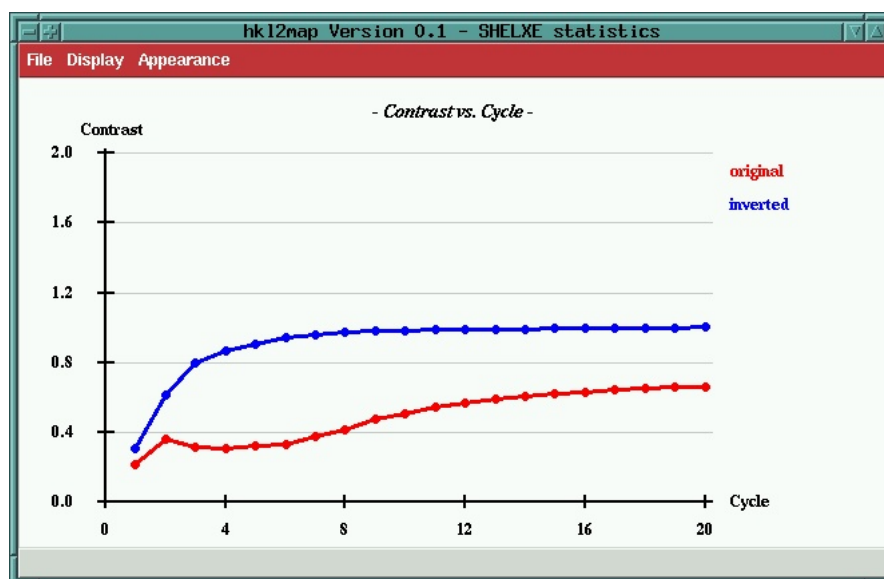

Figure 8: The plot of contrast vs. cycle number. The contrast-values for the original and the inverted substructure are shown in red and blue respectively. Here, the inverted substructure is the correct one.

When a clear difference between contrasts for the two enantiomorphs is seen, the phasing job using the wrong enantiomorph can be interrupted by pressing the respective Finish-Original/Inverted-Button to free the computing resources.

3.5 Displaying Electron Density

The electron density maps corresponding to the phase sets produced by SHELXE can be inspected by pressing the Run-Xfit-Button corresponding to the required substructure enantiomorph in the SHELXE-panel of HKL2MAP.

4 Acknowledgements

References

[1] J. Li, U. Derewenda, Z. Dauter, S. Smith, and Z.S. Derewenda. Crystal structure of the escherichia coli thioesterase ii, a homolog of the human nef binding enzyme. *Nat.Stru.Biol.*, 7:555–559, 2000.
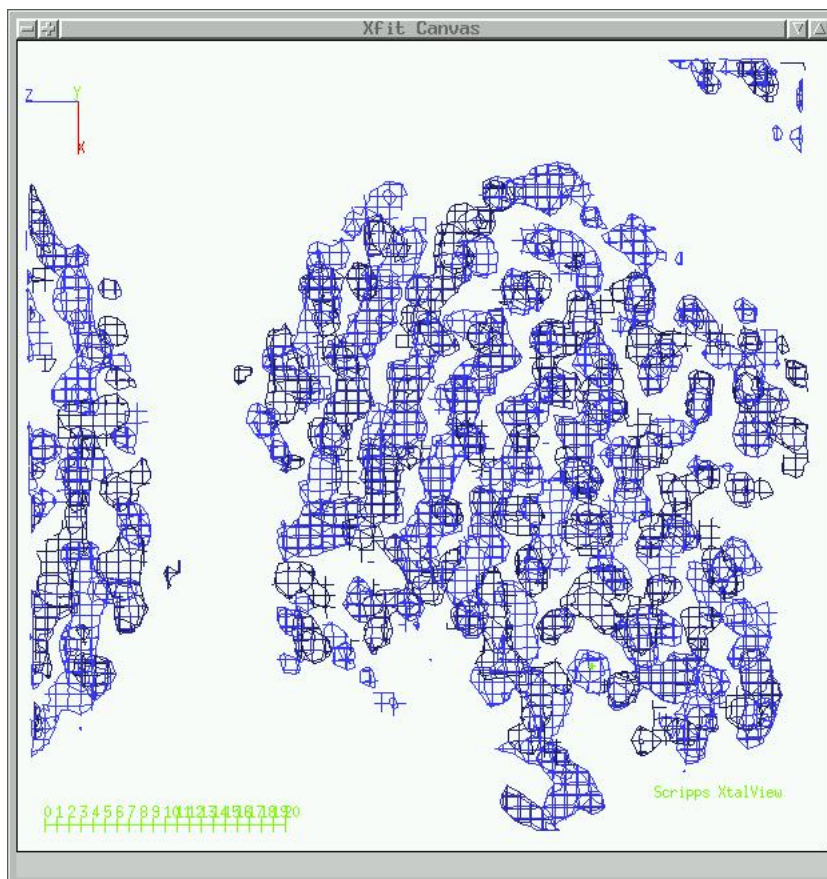
Figure 9: Electron density map for TE-II as displayed by XFIT.