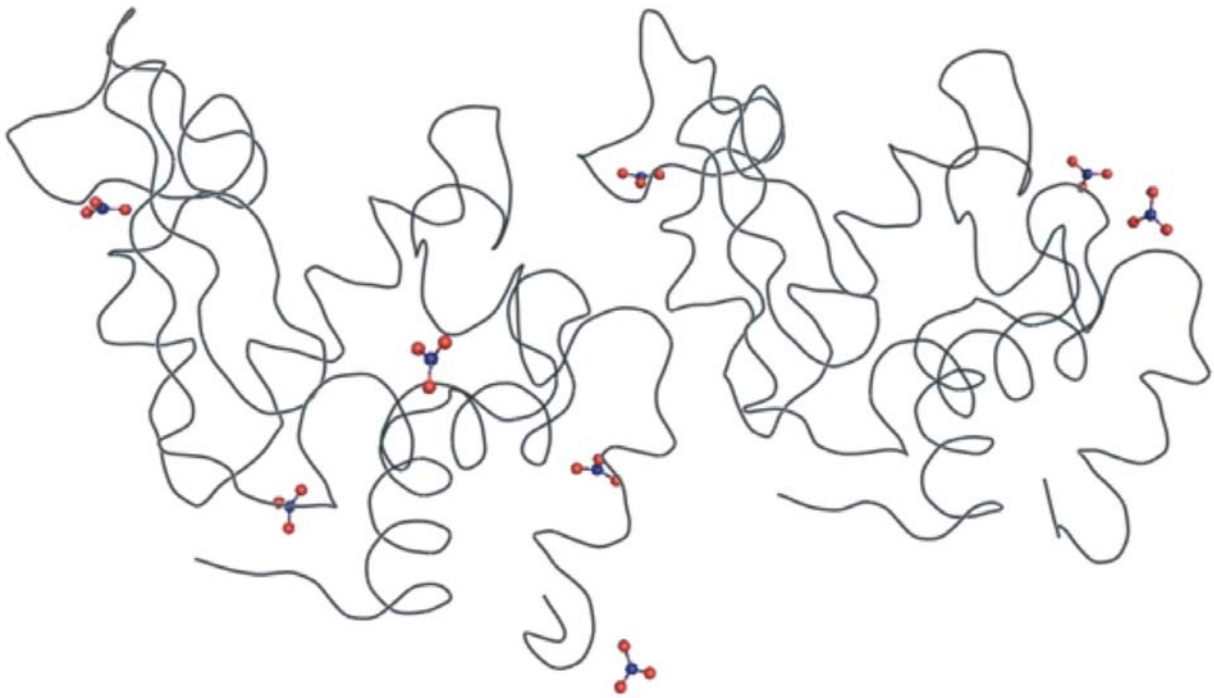
A tutorial for learning and teaching macromolecular crystallography

Annette Faust, Santosh Panjikar, Uwe Mueller, Venkataraman Parthasarathy, Andrea Schmidt, Victor S. Lamzin and Manfred S. Weiss



Reference: Faust *et al*. (2008). J. Appl. Cryst. (in press).

Experiment 3: Molecular Replacement on monoclinic Lysozyme

Lysozyme is a 129 amino acid enzyme that dissolves bacterial cell walls by catalyzing the hydrolysis of 1,4-β-linkages between N-acetylmuramic acid and N-acetyl-D-glucosamine residues in the peptidoglycan layer and between N-acetyl-D-glucosamine residues in chitodextrins. It is abundant in a number of secreted fluids, such as tears, saliva and mucus. Lysozyme is also present in cytoplasmic granules of the polymorphonuclear neutrophils (Voet *et al.*, 2006). Large amounts of lysozyme can also be found for instance in egg whites. The crystal structure of hen egg-white lysozyme (HEWL) based on crystals belonging to the tetragonal space group $P4_32_12$, was the first enzyme structure published (Blake *et al.*, 1965). Over the years, HEWL has been crystallized in many different crystal forms (for an overview see Brinkmann *et al*., 2006) and has become a standard object for methods developments but also for teaching purposes.

In this experiment, the structure of monoclinic HEWL is determined by Molecular Replacement (MR) using the structure of tetragonal HEWL as a search model. MR is a method to determine a structure in cases where a similar structure is already known. If the similar structure can be correctly oriented and positioned in the unit cell of the structure to be solved, it can be used as a starting point for phase calculation and refinement. Currently, about two thirds of all new structures deposited with the PDB (Berman *et al*., 2000) are solved using MR (Long *et al*., 2008).

```
         10         20         30         40         50         60         70
          |          |          |          |          |          |          |
KVFGRCELAAAMKRHGLDNYRGYSLGNWVCAAKFESNFNTQATNRNTDGSTDYGILQINSRWWCNDGRTP

         80         90        100        110        120      129
          |          |          |          |          |        |
GSRNLCNIPCSALLSSDITASVNCAKKIVSDGNGMNAWVAWRNRCKGTDVQAWIRGCRL
```

Figure 1: Amino acid sequence of hen egg-white lysozyme

1 Crystallisation

Chemicals: hen egg-white lysozyme (M~14600 g/mol, Fluka cat. no. 62970)
CH_3COONa (M=82.03 g/mol, Sigma cat. no. S2889)
CH_3COOH (M=60.0 g/mol, Sigma cat.no. 537020)
$NaNO_3$ (M=84.99 g/mol, Sigma cat. no. S5506)
Milli-Q water
Paraffin oil (Fluka cat. no. 76235)

Monoclinic HEWL (M_r = 14.6 kDa, Fluka cat. no. 62970) crystals (Figure 3c) were prepared according to a recipe described before (Saraswathi *et al*., 2002) by mixing 12 µl of protein solution (20 mg/ml lysozyme in 50 mM sodium acetate pH 4.5) and 12 µl of reservoir solution containing 50 mM sodium acetate pH 4.5 and 4% sodium nitrate and equilibrating the drop against the reservoir. Crystals belonging to the monoclinic space group $P2_1$ with unit-cell parameter a = 27.4 Å, b = 62.3 Å, c = 59.5 Å and β = 90.5º grew within a few days. They were cryo-protected using paraffin oil and usually diffracted X-rays to better than 1.3 Å.

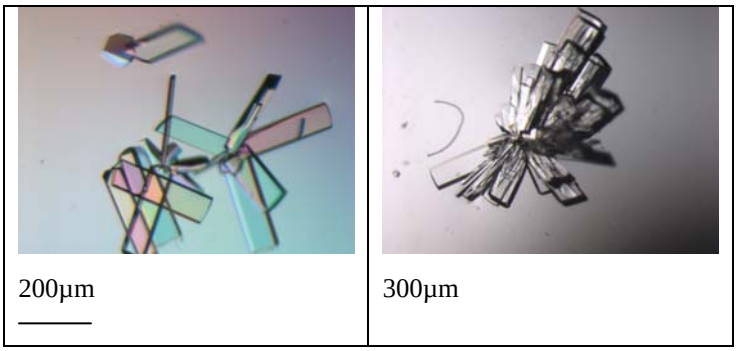


Figure 2: Monoclinic HEWL crystals

2 Data Collection

X-ray diffraction data have been collected at beam line BL 14.2 at the BESSY synchrotron in Berlin Adlershof. The beam line is equipped with a MARCCD detector (165mm) from MARRESEARCH (Norderstedt, Germany) and a MARdtb goniostat (MARRESEARCH, Norderstedt, Germany).

The relevant data collection parameters are given below:

wavelength	0.9 Å
detector distance:	100 mm
oscillation range/image:	1.0º
no. of images:	360
exposure time/image:	3.4 sec
path to images:	experiment3/data
image names:	lys_mono_low _1_###.img

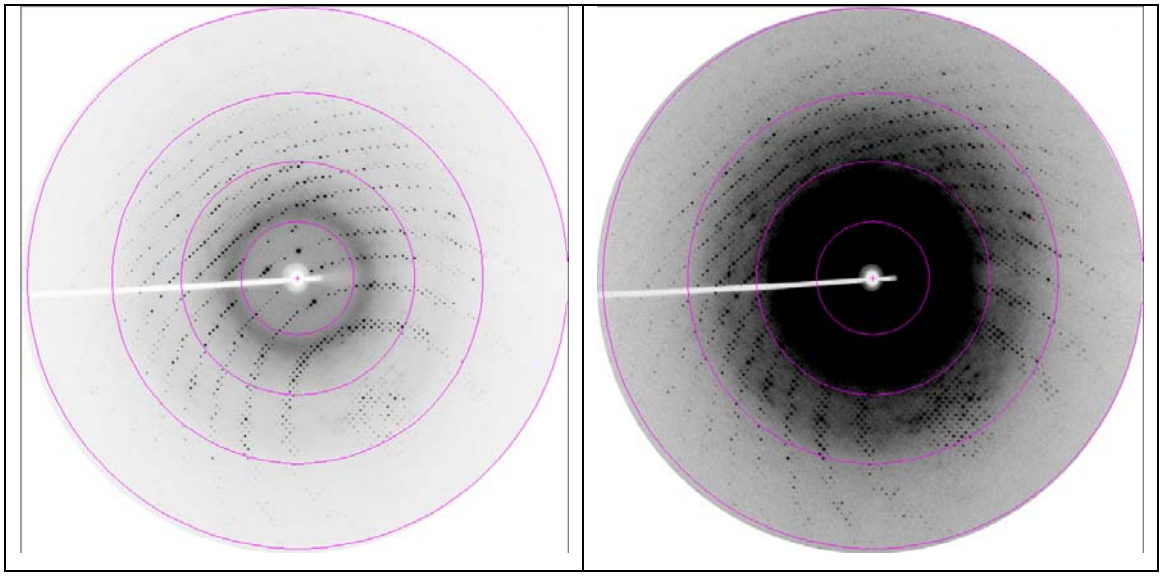


Figure 3: Diffraction image of a crystal of monoclinic HEWL displayed using different contrast levels. The resolution rings are shown at 5.4, 2.7, 1.8 and 1.4Å, respectively.

3 Data Processing

The data were indexed, integrated and scaled using the program XDS (Kabsch, 1993). XDS is able to recognize compressed images, therefore it is not necessary to unzip the data before using XDS. (For use with other programs this will be necessary and can be done using the command **bunzip2** *.bz2). XDS needs only one input file. This has to be called XDS.INP, no other name will be recognized by the program. In XDS.INP the image name given must not include the zipping-format extension (*.img instead of *.img.bz2). Further, XDS has a very limited string length (80) to describe the path to the images. Therefore it may be necessary to create a soft link to the directory containing the images by using the command **ln -s** /path/to/images/ ./images. The path to the images in XDS.INP will then be ./images/.

- **indexing** **1st run of XDS**

Before running XDS, the XDS.INP file has to be edited such it contains the correct data collection parameters. To estimate the space group and the cell parameters the space group number in XDS.INP has to be set to 0. These parameters will be obtained in the output file IDXREF.LP.

JOBS= XYCORR INIT COLSPOT IDXREF space group number=0

XYCORR computes a table of spatial correction values for each pixel.

INIT determines an initial background for each detector pixel and finds the trusted region of the detector surface.

COLSPOT collects strong diffraction spots from a specified subset of the data images.

IDXREF interprets observed spots by a crystal lattice and refines all diffraction parameters.

The IDXREF.LP output file contains the results of the indexing. For monoclinic HEWL the correct space group is $P2_1$ (space group number 4) with unit cell parameters a=27.40 Å, b=62.30 Å, c=59.50 Å and β=90.50.

- **integration** **2nd run of XDS**

After determination of space group and cell parameters all images will be integrated and corrections for radiation damage, absorption, detector etc. will be calculated in a second XDS run.

DEFPIX defines the trusted region of the detector, recognizes and removes shaded areas, and eliminates regions outside the resolution range defined by the user.

XPLAN helps planning data collection. Typically, one or a few data images are collected initially and processed by XDS. XPLAN reports the completeness of data that could be expected for various starting angles and total crystal rotation.

Warning: If data were initially processed for a crystal with unknown cell constants and space group, the reported results will refer to space group P1.

INTEGRATE collects 3-dimensional profiles of all reflections occurring in the data images and estimates their intensities

CORRECT corrects intensities for decay, absorption and variations of detector surface sensitivity, reports statistics of the collected data set and refines the diffraction parameters using all observed spots.

The file CORRECT.LP contains the statistics for the complete data set after integration and corrections. After truncation a file named XDS_ASCII.HKL will be written out, which contains the integrated and scaled reflections. If the cell parameters and the space group are known already one can run XDS with JOBS=ALL.

- **scaling run XSCALE**

The collected reflections on the different images have to be put on a common scale. The correction factors are determined and applied to compensate absorption effects and radiation damage. Individual reflections can be corrected for radiation damage (0-dose corrections). XSCALE writes out a *.ahkl file, which can be converted with XDSCONV to be used within the CCP4-suite (Collaborative Computational Project, 1994) or other programs.

Table 1: Data processing statistics (from XSCALE.LP).

Resolution limits [Å]	10.0 - 1.60 (1.70-1.60)
Unit cell parameters, a, b, c,β [Å,°]	27.40, 62.30, 59.50, 90.50
Space group	$P2_1$
Mosaicity [˚]	0.5
Total number of reflections	317,620
Redundancy	7.3 (5.5)
Unique reflections	43,663
Completeness [%]	99.4 (94.5)
I/σ(I)	27.4 (4.7)
$R_{r.i.m.}$ / R_{meas} [%]	4.2 (40.2)
Wilson B-factor [Å^2]	18.9

- **converting *.ahkl to *.mtz run XDSCONV with XDSCONV.INP**

XDSCONV.INP:

```
OUTPUT_FILE=lys_mono_xds.mtz CCP4
INPUT_FILE=lys_mono.ahkl
```

XDSCONV creates an input file F2MTZ.INP for the final conversion to binary mtz-format. To run the CCP4 programs F2MTZ and CAD, just type the two commands:

```
f2mtz HKLOUT temp.mtz < F2MTZ.INP
cad HKLIN1 temp.mtz HKLOUT lys_mono_ccp4.mtz << EOF
LABIN FILE 1 ALL
END
EOF
```

Alternatively, the XDS_ASCII.HKL file can be converted with COMBAT (to mtz-format) and this mtz-file can be used as an input file for SCALA in CCP4.

4 Structure Solution

The structure can be solved using the SAD-protocol (run in the advanced version) of AUTO-RICKSHAW: the EMBL-Hamburg automated crystal structure determination platform (Panjikar *et al.*, 2005). AUTO-RICKSHAW can be accessed from outside EMBL under www.embl-hamburg.de/AutoRickshaw/LICENSE (a free registration may be required, please follow the instructions on the web page). In the following the automatically generated summary of AUTO-RICKSHAW is printed together with the results of the structure determination:

The structure was solved using the MR-protocol of Auto-Rickshaw with tetragonal HEWL (PDB entry 193L, Vaney *et al.*, 1996) as a starting model. The input diffraction data (file XDS_ASCII.HKL) were uploaded and then prepared and converted using programs of the CCP4-suite. The molecular replacement step was done using MOLREP (Vagin and Teplyakov, 1997) with a resolution cut-off of 4 Å to find the two molecules in the asymmetric unit. Despite a very high initial R-factor of 73% (correlation coefficient 43%), the solution was correct as was demonstrated by subsequent refinement. This was performed to a resolution of 3.0 Å using the program CNS (Bruenger *et al.*, 1998) in four consecutive steps: rigid body refinement, a minimisation step, B-factor refinement and a second minimisation step. At this point the R- and R_{free}-values were 24.9 and 33.5%, respectively. Further refinement was done in REFMAC5 using all available data to R- and R_{free}-values of 28.3 and 31.5%. The model was completed and further modified using COOT and refined using REFMAC5. Figure 6 shows the final electron density with some nitrate ions clearly visible. For more detailed information see the AUTO-RICKSHAW output (directory experiment3/autorickshaw). Figure 3 shows the cartoon representation of the two molecules in the asymmetric unit. Clear electron density can be found where the nitrate ions are bound (see Figure 4).

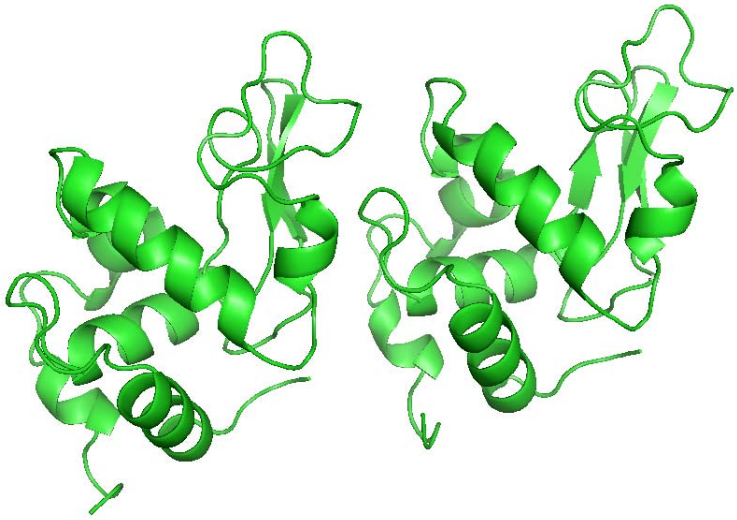


Figure 3: Cartoon representation of the two molecules of HEWL in the asymmetric unit.

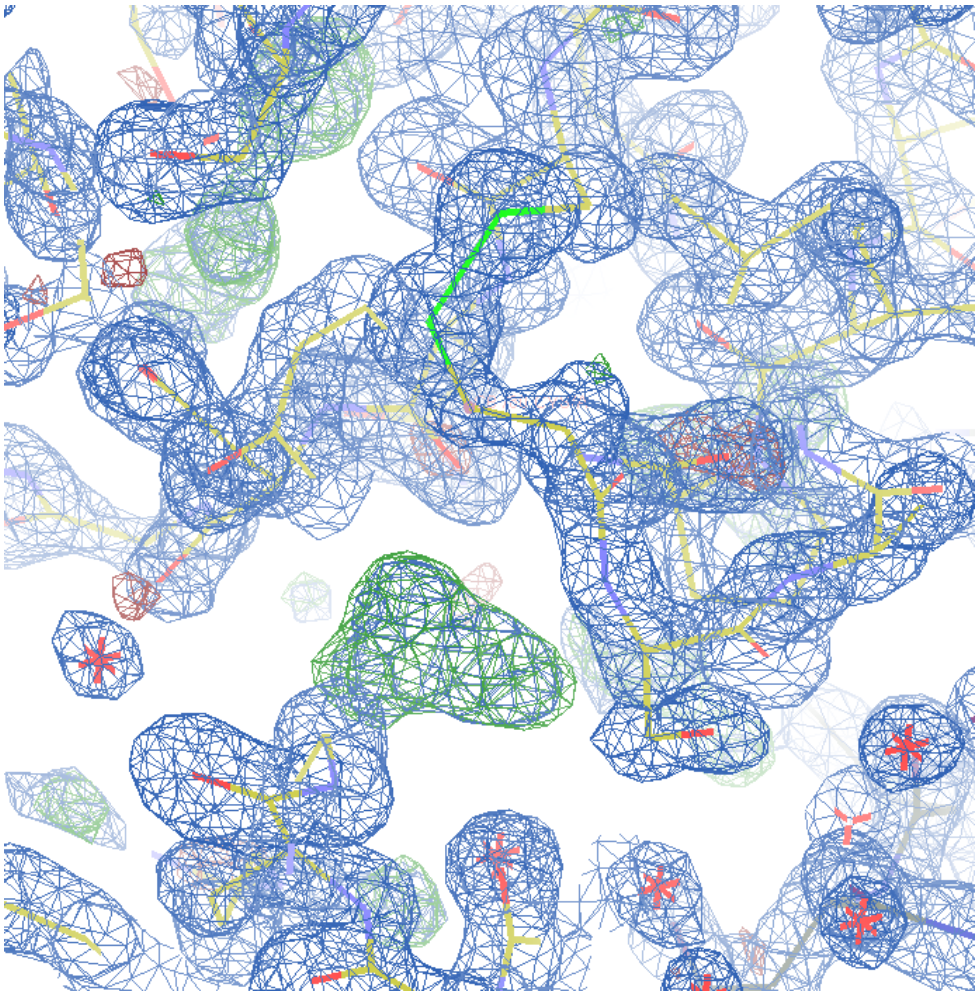


Figure 4: Experimental electron density map showing the bound nitrate ions. The ($2F_{obs}$-F_{calc},α_{calc})-map (blue) is contoured at 1.2 σ, the (F_{obs}-F_{calc},α_{calc})-map (green and red) at +3.0 and – 3.0 σ, respectively.

5 References

Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N. &Bourne, P. E. (2000). *Nucleic Acids Research* **28**, 235-242.

Blake, C. C. F., Koenig, D. F., Mair, G. A., North, A. C. T., Philipps, D. C. & Sarma, V. R. (1965). *Nature* **206**, 757-761.

Brinkmann, C., Weiss, M. S. & Weckert, E. (2006). *Acta Cryst*. **D62**, 349-355.

Bruenger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Gross-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T. & Warren, G. L. (1998). *Acta Cryst.* **D54**, 905-921.

Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* **D50**, 760-763.

Emsley, P. & Cowtan, K. (2004). *Acta Cryst.* **D60**, 2126-2132.

Kabsch, W. (1993*). J. Appl. Cryst.,* **26**, 795-800.

Long, F., Vagin, A. A., Young, P. & Murshudov, G. N. (2008). *Acta Cryst.* **D64**, 125-132.

Morris, R. J., Perrakis, A. & Lamzin, V. S. (2002). *Acta Cryst.* **D58**, 968-975.

Mueller-Dieckmann, C., Panjikar, S., Schmidt, A., Mueller, S., Kuper, J., Geerlof, A., Wilmanns, M., Singh, R. K., Tucker, P. A. &Weiss, M. S. (2007). *Acta Cryst.* **D63**, 366-380.

Murshudov, G. N., Vagin, A. A. & Dodson, E. J. (1997). *Acta Cryst.* **D53**, 240-255.

Panjikar, S., Parthasarathy, V., Lamzin, V. S., Weiss, M. S. & Tucker, P. A. (2005). *Acta Cryst.* **D61**, 449-457.

Perrakis, A., Morris, R. J. & Lamzin, V. S. (1999). *Nature Struct. Biol.* **6**, 458-463.

Saraswathi, N. T., Sankaranarayanan, R. & Vijayan, M. (2002). *Acta Cryst.* **D58**, 1162-1167.

Vagin, A., Teplyakov, A. (1997). *J. Appl. Cryst.,* **30**, 1022-1025.

Vaney, M. C., Maignan, S., Ries-Kautt, M. & Ducruix, A. (1996). *Acta Cryst.* **D52**, 505-517.

Voet, D., Voet, J. & Pratt, C. W. (2006). Fundamentals in Biochemistry - Life at the molecular level, 2nd Edition, John Wiley & Sons, Inc., Hoboken, NJ, USA.