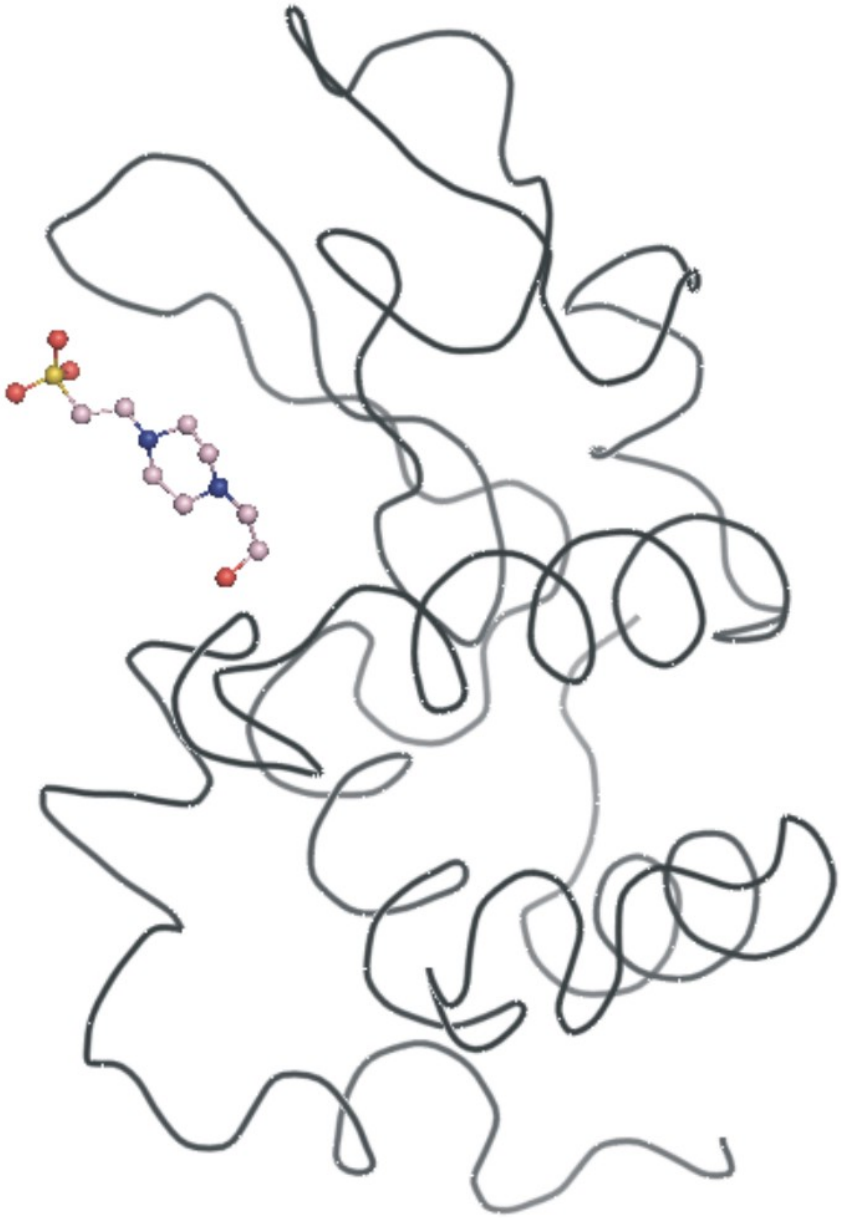
A tutorial for learning and teaching macromolecular crystallography

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 5: Identification of a bound ligand in the active site of tetragonal 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.

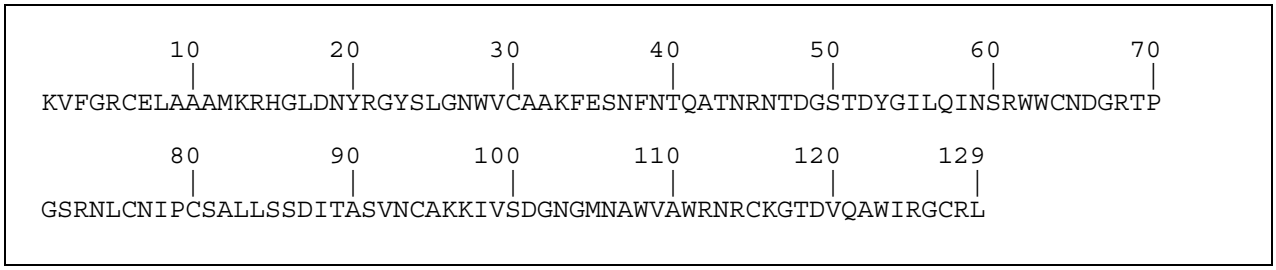


Figure 1: Amino acid sequence of HEWL

The knowledge of the three-dimensional structure of a drug target is the first step in crystallographic structure-based drug design. The architecture of the active site of the target molecule can be used to design potential molecules, which can bind to the active site and may inhibit the enzymatic reaction of the target molecule. The use of X-ray crystallography nowadays allows the fast and high-throughput screening of a large number of chemically different ligands, which may become potential lead compounds. For this, advantage can be taken of the recent approaches towards automatisation and high-throughput in crystallization and data collection. Data collected from crystals soaked with a ligand can be quickly evaluated using the direct refinement approach (by skipping the molecular replacement step). The method is also more generally applicable to identify inhibitors, substrate or products bound to a macromolecule in order to evaluate reaction mechanisms and other things.

If tetragonal HEWL crystals are grown in the presence of HEPES and MPD according to a published procedure (Weiss *et al.*, 2000), a bound HEPES molecule can be identified from the

difference density map and the position can be verified and refined using the anomalous signal of the sulfur atom.

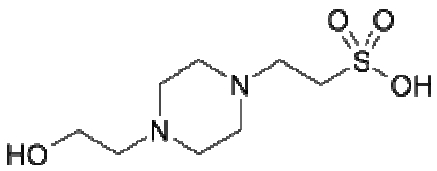


Figure 2: Chemical structural of HEPES

1 Crystallisation

Chemicals: hen egg-white lysozyme (M= 14.7 kDa, Fluka cat. no. 62970)
HEPES (N-(2-Hydroxyethyl)piperazine-N-(2-ethanesulfonic acid) (M= 238.3 g/mol, Sigma cat. no H3375)
sodium hydroxide (M= 40.0 g/mol, Sigma cat. no. S8045)
MPD (2-Methyl-2,4-pentanediol) (M= 118.17 g/mol, Sigma cat. no. M9671)
Milli-Q water

Another tetragonal crystal form of HEWL was grown by mixing 4 µl of protein solution (30 mg/ml in water) and 4 µl of reservoir solution containing 50-100 mM HEPES pH 7.2 and 65-70% (v/v) MPD and equilibrating the drop against the reservoir. The crystals (space group $P4_32_12$, space group number 96) appeared within a few days and exhibited the unit-cell parameters of a = 78.7 and c = 37.1 Å. For flash cooling to 100 K they were mounted directly from the drop. They typically diffracted X-rays to a resolution of better than 1.6 Å.

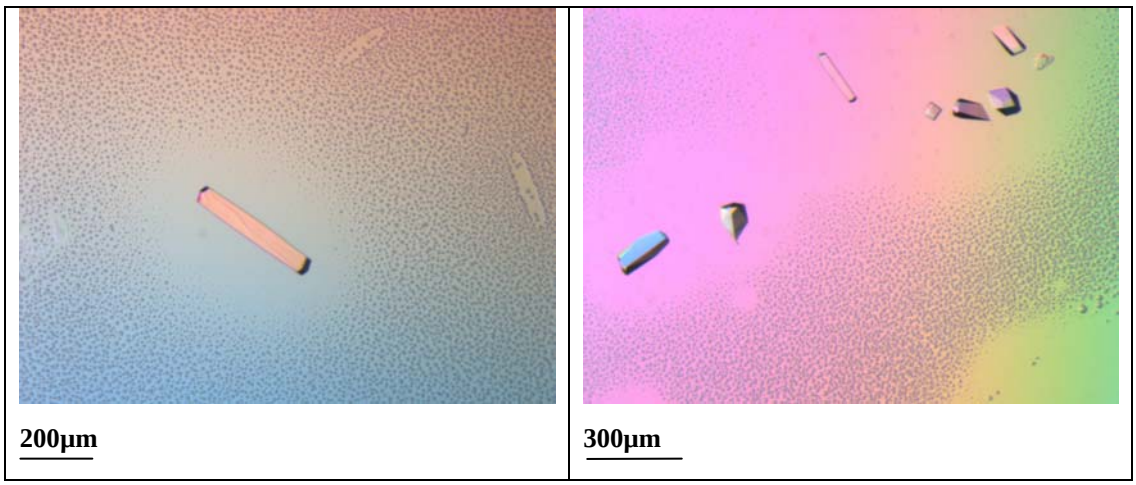


Figure 3: Tetragonal HEWL crystals grown in the presence of HEPES and MPD

2 Data Collection

X-ray diffraction data has been collected at the tunable beam line BL 14.1 at the BESSY synchrotron in Berlin-Adlershof. The beam line is equipped with a MARMosaic CCD detector (225mm) from MARRESEARCH (Norderstedt, Germany) and a MARdtb goniostat (MARRESEARCH, Norderstedt, Germany).

The relevant data collection parameters are given below:

wavelength	1.70 Å
detector distance:	70mm
oscillation range/image:	1.0º
no of images:	180
exposure time/image:	7.4 sec
path to images:	experiment5/data
image names:	lys_hepes _1_###.img

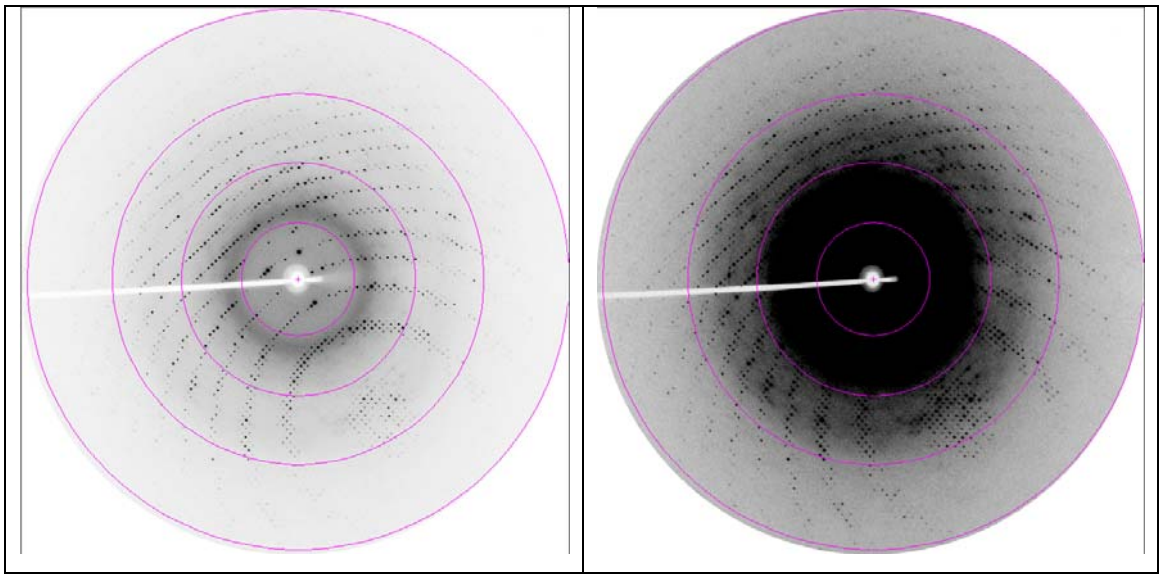


Figure 4: Diffraction image of tetragonal lysozyme displayed at different contrast levels. The resolution rings are at 7.0, 3.5, 2.3 and 1.8 Å, respectively.

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 so that 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 tetragonal HEWL, the correct space group is $P4_32_12$ (space group number 96) with unit cell parameters of a=78.7 and c=37.1 Å.

- **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 data set will 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 has to be converted using XDSCONV to be use within the CCP4-suite (Collaborative Computational Project 1994) or other programs.

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

Resolution limits [Å]	10.0 -2.00 (2.10-2.00)
Unit cell parameters a, c [Å]	78.7, 37.1
Space group	$P4_32_12$
Mosaicity [˚]	0.45
Total number of reflections	107512
Unique reflections	15043
Redundancy	7.1 (6.8)
Completeness [%]	99.9 (100.0)
I/σ(I)	24.5 (3.9)
$R_{r.i.m.}$ / R_{meas} [%]	6.4 (72.2)
Wilson B-factor [Å^2]	31.4

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

XDSCONV.INP:

```
OUTPUT_FILE=lys_hepes_xds.mtz CCP4
INPUT_FILE=lys_hepes.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_hepes_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 the CCP4-package (Collaborative Computational Project, 1994).

4 Structure Solution

The structure can be solved using the MR-protocol of Auto-Rickshaw: the EMBL-Hamburg automated crystal structure determination platform (Panjikar *et al.*, 2005) with tetragonal lysozyme (PDB entry 1DPW, Weiss *et al.*, 1996) as a starting model. 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 input diffraction data (file XDS_ASCII.HKL) was uploaded and the prepared and converted for use in Auto-Rickshaw using programs of the CCP4-suite (Collaborative Computational Project, 1994). Because the cell parameters of model and target were the same Auto-Rickshaw skipped the MOLREP (Vagin and Teplyakov, 1997) run. The refinement was performed to a resolution cut-off of 3.0 Å in CNS (Bruenger *et al.*, 1998) in four consecutive steps: a rigid body refinement, a minimisation step, a B-factor refinement and a second minimisation step. Further refinement was then done in REFMAC5 (Murshudov *et al.*, 1997) using all available data. As the R-factor after REFMAC5 refinement was still above 30% the model was rebuild using ARP/wARP (Perrakis *et al.*, 1999; Morris *et al.*, 2002). The ARP/wARP model contained 126 out of 129 amino acids. The model was completed and further refined using COOT (Emsley and Cowtan, 2004) and REFMAC5. Clear electron density was found for a bound HEPES molecule (Figure 5). To verify the position of the ligand, an anomalous difference Fourier electron density map was calculated to identify the sulfur atom (Figure 6) of the HEPES molecules.

Alternatively, structure solution can directly proceed from a model of tetragonal lysozyme, such as for instance PDB entry 1DPW (Weiss *et al.*, 1996). After some rounds of rigid body refinement, structure refinement can be continued.

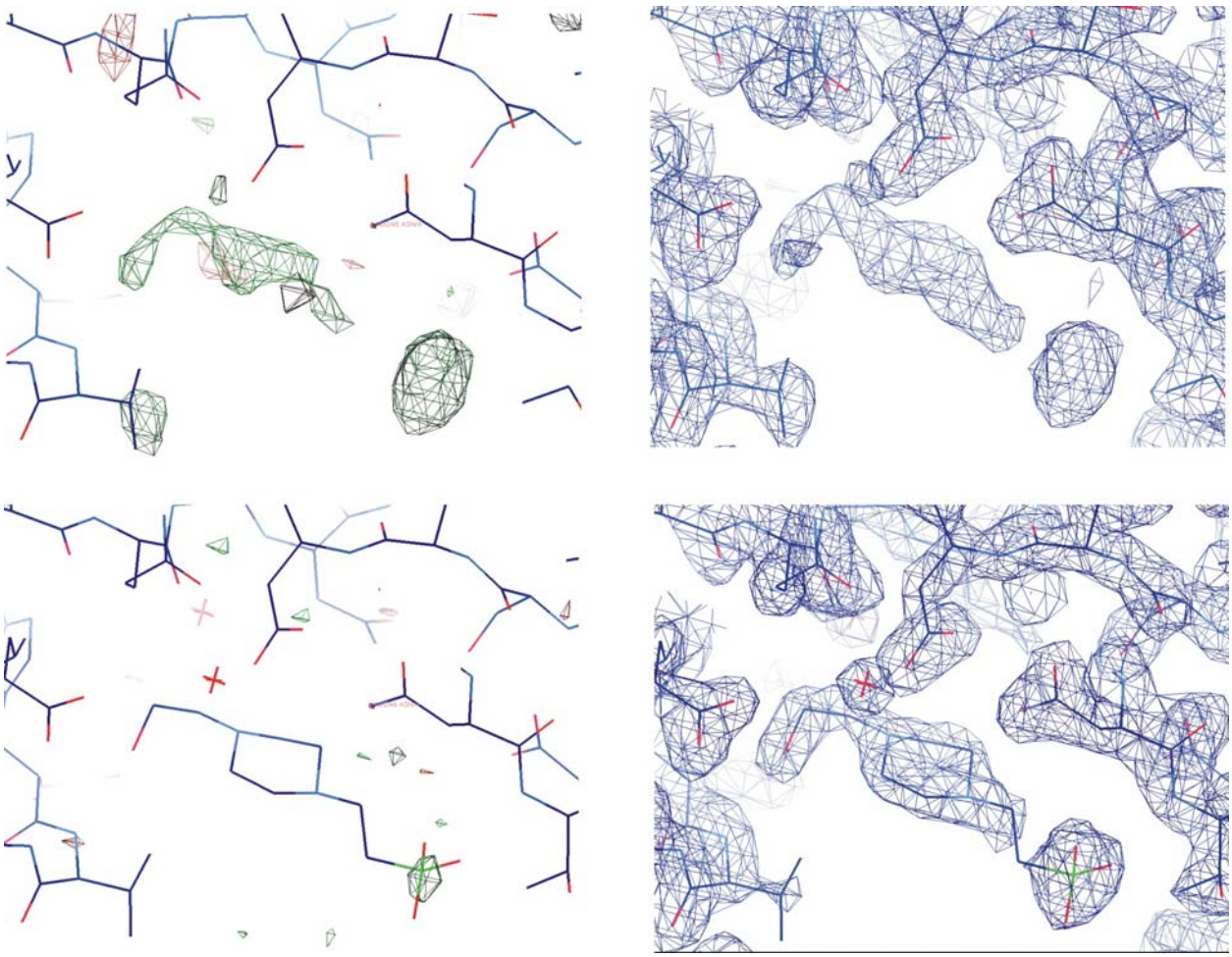


Figure 5: HEPES binding to HEWL. Top left panel: (F_{obs}-F_{calc}, α_{calc})-electron density contoured at +2.0 σ (green) and -2.0 σ (red), top right panel ($2F_{obs}$-F_{calc},α_{calc})-electron density contoured at 1.0 σ. In both cases the HEPES model was not included into the phase calculation. Bottom left panel: Anomalous difference Fourier electron density map contoured at +3.0 σ showing a peak at the position of the S-atom of the HEPES molecule lending further support to the identification of the density as belonging to HEPES. Bottom right panel: Final ($2F_{obs}$-F_{calc}, α_{calc})-electron density contoured at 1.5 σ for the final model including the HEPES ligand.

5 References

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.

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.

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.

Weiss, M. S., Palm, G. J. & Hilgenfeld, R. (2000). *Acta Cryst.* **D56**, 952-958.