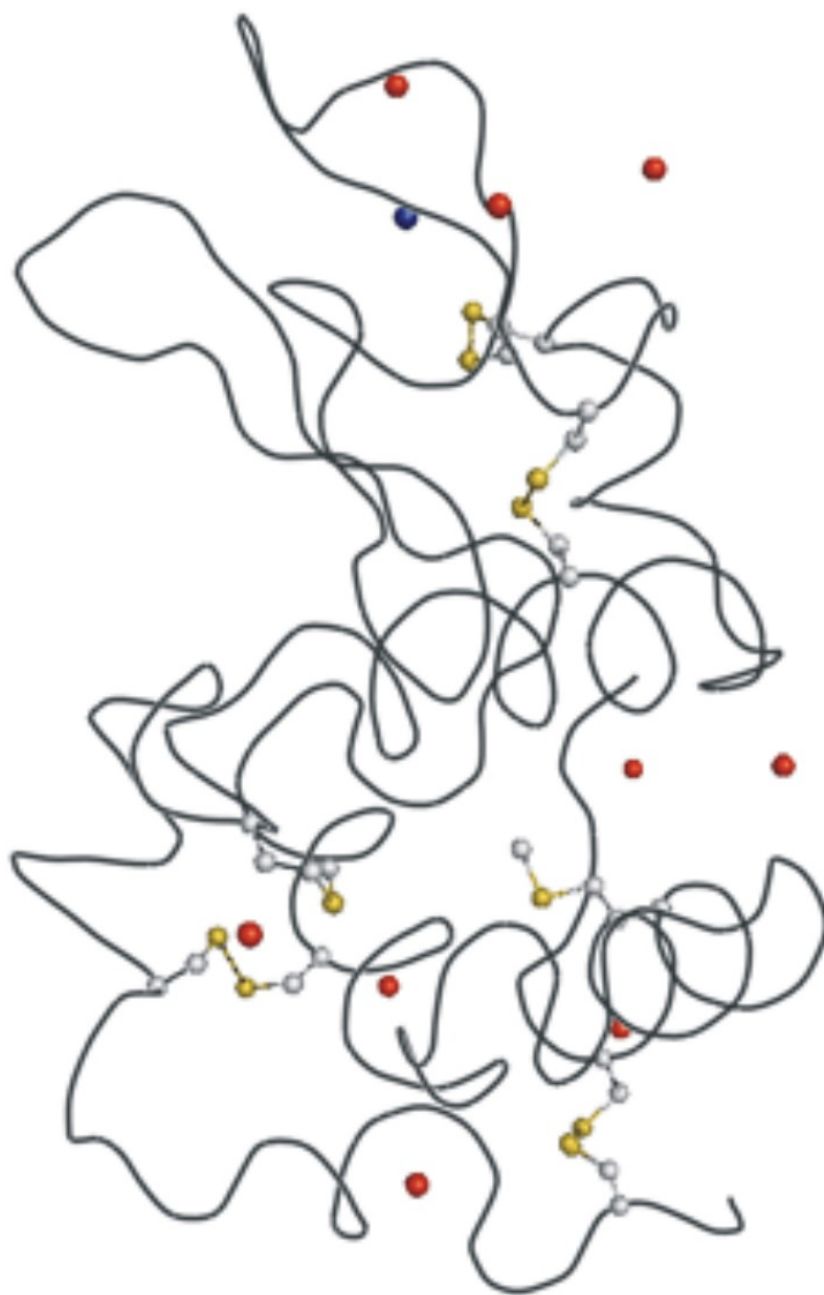# A tutorial for learning and teaching macromolecular crystallography

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

Experiment 4: Identification of bound solvent ions in 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.

```
         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

The 01 August 2006 release of the Protein Data Bank (Berman *et al.*, 2000) contained 34,776 macromolecular crystal structures determined by X-ray crystallographic methods. A search for prosthetic groups/ligands in these structures yielded the result that 1066 of them (3.1%) were reported to contain one or more phosphate ions (PDB ID PO4), 4050 (11.6%) sulfate ions (PDB ID SO4), 1963 (5.6%) chloride ions (PDB ID CL), 848 (2.4%) potassium ions (PDB ID K) and 3295 (9.5%) calcium ions (PDB ID CA). These numbers are surprisingly low given the facts that NaCl or KCl are part of almost every buffer used in protein crystallization and that ammonium sulfate is one of the most widely used precipitants in protein crystallization. A possible explanation for this may be that partially occupied or partially disordered light-atom-containing ions are often mistakenly interpreted as water molecules. Since most protein structure refinements are based on a diffraction data set collected at a rather short wavelength of about 1.0 Å, additional information in the form of anomalous differences originating from these atoms is typically not available. Furthermore, diffraction data sets collected for refinement purposes are

typically collected with the strategy of obtaining an as complete as possible set of structure-factor amplitudes to the diffraction limit of the crystal. Comparatively little effort is spent on collecting good and complete anomalous differences in such experiments. As can be seen from Figure 1, the anomalous scattering lengths of elements 11-20 are about three to four times larger at a wavelength of 2.0 Å than at 1.0 Å. Thus, it has recently been suggested that complementing a short-wavelength diffraction data set, which is usually used for refinement, with a long-wavelength data set can provide sufficient additional information to unequivocally identify the complete anomalously scattering substructure (Djinovic Carugo *et al*., 2005; Mueller-Dieckmann *et al.*, 2005). In this experiment, excellent anomalous differences will be collected at a wavelength close to the copper K_α-absorption edge, with the aim of identifying the complete anomalously scattering substructure in tetragonal HEWL.
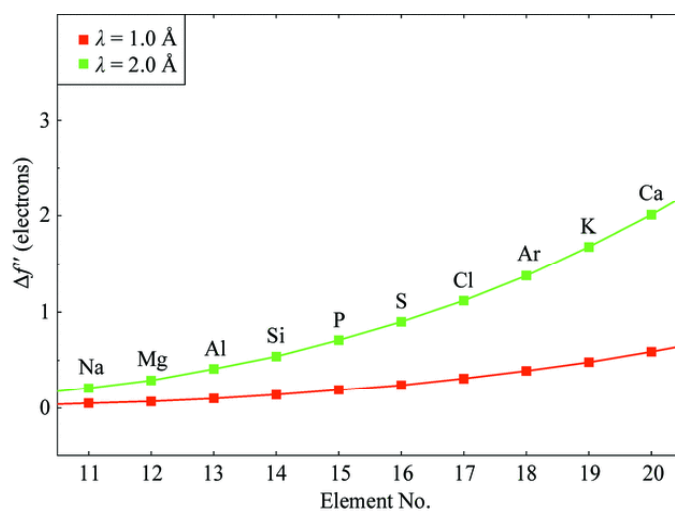

Figure 2: Anomalous scattering length Δf" values in units of electrons at λ = 1.0 Å (red) and λ = 2.0 Å (green) for elements 11-20 according to Cromer & Liberman (1970).

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)

                      NaCl (M=58.44 g/mol, Sigma cat. no. S7653)

                      Milli-Q water

                      Paraffin oil (Fluka cat. no. 76235)

Tetragonal crystals of HEWL were grown as described by Weiss *et al*. (2000) by mixing 4 µl of protein solution (30 mg/ml in water) and 4 µl of reservoir solution containing 50 mM sodium acetate pH 4.6 and 5% (w/v) sodium chloride and equilibrating the drop against the reservoir. The crystals belonging to space group $P4_32_12$ (space group number 96) and exhibiting the usual unit-cell parameters of a = 78.8 Å and c = 37.2 Å appeared within few days (Figure 3). Prior to flash cooling at 100 K they were transferred into paraffin oil. They typically diffracted X-rays to a resolution better than 1.6 Å.
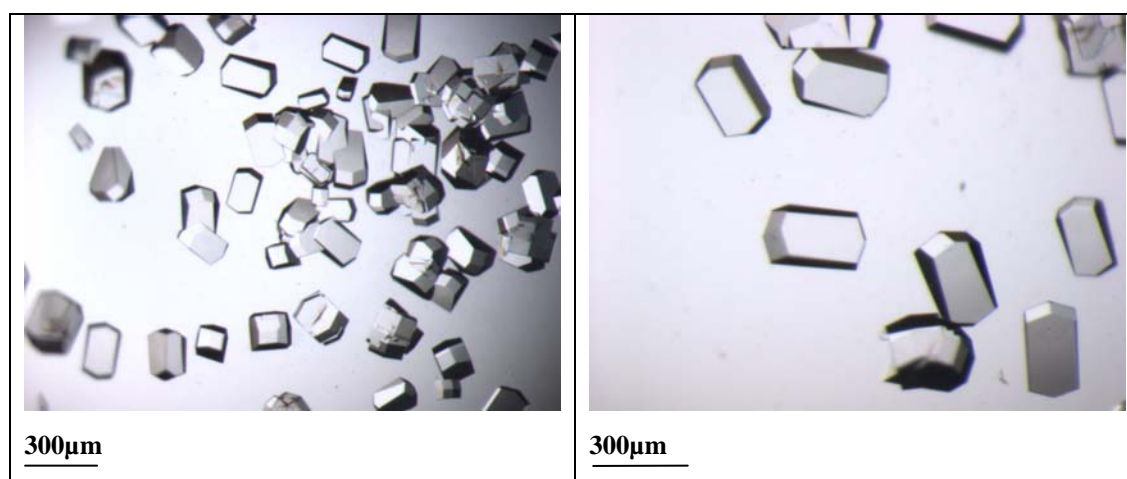

Figure 3: Tetragonal HEWL crystals.

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:

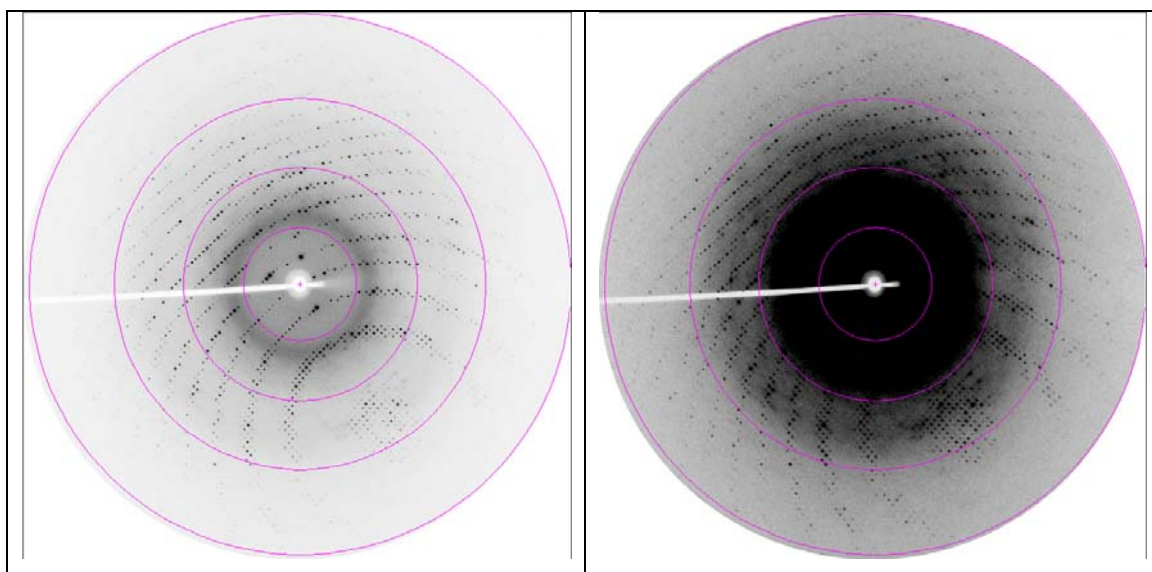|	
wavelength	1.5418 Å
detector distance:	100 mm
oscillation range/image:	1.0º
no of images:	360
exposure time/image:	3.4 sec
path to images:	experiment4/data
image names:	lys_cl _1_###.img

Figure 4: Diffraction image of tetragonal lysozyme displayed at different contrast levels. The resolution rings shown are at 7.5, 3.8, 2.5 and 1.9 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 <u>must not include</u> 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 lysozyme, the correct space group is P4$_3$2$_1$2 (space group number 96) with unit cell parameters of a=78.8 and c=37.2 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.

<u>Warning</u>: 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 CORRECT.LP file contains the statistics for the complete data set after integration and corrections. After truncation, a file name 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 (CCP4, 1994) or other programs.

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

Resolution limits [Å]	10.0 - 1.80 (2.00 - 1.80)
Unit cell parameters a, c [Å]	78.8, 37.2
Space group	$P4_32_12$
Mosaicity [°]	0.2
Total number of reflections	285616
Unique reflections	20633
Redundancy	13.9 (11.7)
Completeness [%]	99.3 (97.7)
I/σ(I)	43.0(17.3)
$R_{r.i.m.}$ / R_{meas} [%]	4.5 (4.3)
Wilson B-factor [$Å^2$]	22.0

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

XDSCONV.INP:

```
OUTPUT_FILE=lys_tetra_xds.mtz CCP4
INPUT_FILE=lys_tetra.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 tetra_lys_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 within the CCP4-package.

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 193L, Vaney *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) were uploaded and then prepared and converted using programs of the CCP4-suite. Because the cell parameters of the model structure and the target structure were very similar, Auto-Rickshaw skipped the MR step and proceeded directly to refinement. The refinement was then performed to a resolution cut-off of 3.0 Å in CNS 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 22.6 and 32.4%, respectively. Further refinement was then done in REFMAC5 using all available data to R- and R_{free}-values of 25.5 and 28.2%. The model was completed using COOT and further refined using REFMAC5. An anomalous difference Fourier electron density map using the final model phases was calculated to identify the bound chlorine and sodium atoms. Table 2 contains the highest peaks found in the anomalous difference Fourier electron density map (threshold 4.5 σ above the mean value of the map). Nine chloride atoms can be identified in addition to the intrinsically present ten sulphur atoms of HEWL. Figure 5 shows the refined structure of HEWL overlaid with the anomalous difference Fourier electron density map identifying the protein sulfur atoms and the surface-bound chlorides ions.

Table 2: List of peaks in the anomalous difference Fourier electron density map ($F^+ - F^-$, $\alpha - 90°$) above 4.0 σ and their classifications.

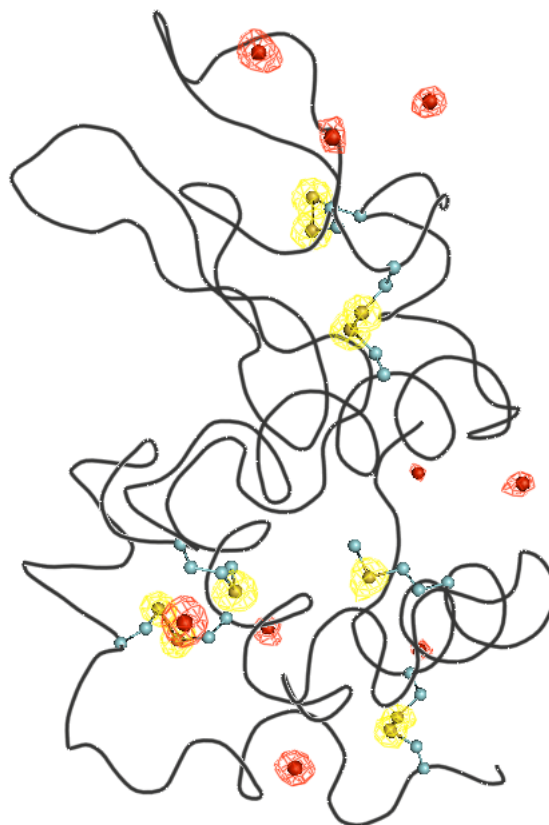| Order	Peak Height/Rms	Atom type
1	15.7	Met105-SD
2	14.8	Cys64-SG
3	14.3	Cys80-SG
4	14.2	Cl-1
5	13.8	Met-12-SD
6	11.5	Cys115-SG
7	11.4	Cl-2
8	11.0	Cys30-SG
9	10.1	Cys76-SG
10	10.1	Cl-3
11	1.1	Cys127-SG
12	9.8	Cys94-SG
13	9.8	Cys6-SG
14	7.4	Cl-4
15	6.8	Cl-5
16	6.1	Cl-6
17	5.9	Cl-7
18	5.4	Cl-8
19	50	Cl-9

Figure 5: Refined structure of tetragonal HEWL presented as a thin tube (in gray). Also shown are the Met and Cys side chains of HEWL (as ball-and-stick) and the surface bound chloride ions (as red spheres). Superimposed on the structure is the anomalous difference Fourier electron density map contoured at 3.0σ (colored in yellow for the peaks occurring at a S-atom and in red for the peaks occurring at a Cl site).

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.

Cromer, D. & Liberman, D. (1970). *J. Chem. Phys*. **53**, 1891-1898.

Djinovic Carugo, K., Helliwell, J. R., Stuhrmann, H. & Weiss, M. S. (2005). *J. Synch. Rad.* **12**, 410-419.

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.