

A tutorial for learning and teaching macromolecular crystallography – version 2010

Annette Faust, Sandra Puehringer, Nora Darowski, Santosh Panjekar,
Venkataraman Parthasarathy, Andrea Schmidt, Victor S. Lamzin, Kay Diederichs,
Uwe Mueller and Manfred S. Weiss



References: A. Faust *et al.* (2008). *J. Appl. Cryst.* **41**, 1161-1172.
A. Faust *et al.* (2010). *J. Appl. Cryst.* **43** (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 P4₃2₁2, 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.

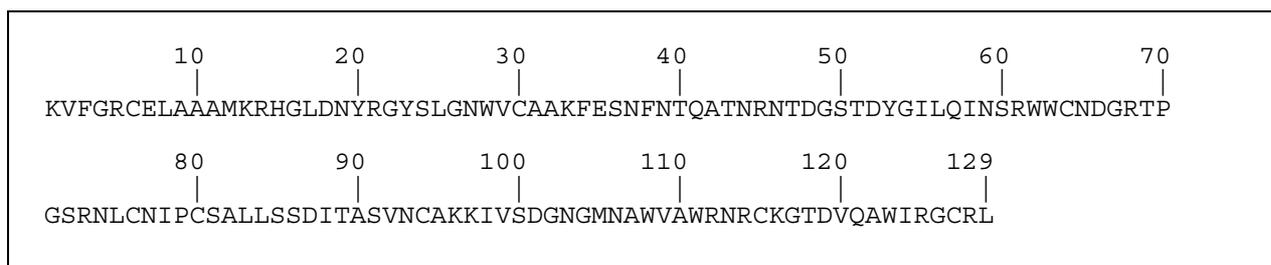


Figure 1: Amino acid sequence of hen egg-white lysozyme

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 automatization 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.

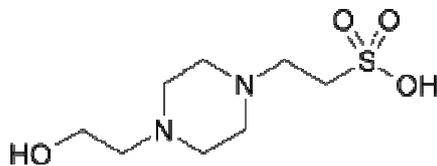


Figure 2: Chemical structural of HEPES

In this experiment, diffraction data will be collected on a crystal of tetragonal HEWL grown in the presence of HEPES. The data will be collected at a wavelength of 1.70 Å. This ensures that the bound ligand can not only be identified in a difference electron density map but that the sulphur atom of the sulfonate group of HEPES is also visible in an anomalous difference electron density map, lending additional proof to the identification of the bound ligand as HEPES.

1 Crystallisation

Chemicals: Hen egg-white lysozyme ($M \approx 14.6$ 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

A tetragonal crystal form of HEWL different from the one of experiment 4 was grown by mixing 4 μ l of protein solution (30 mg/ml in water) and 4 μ l 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 (Figure 3) 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.

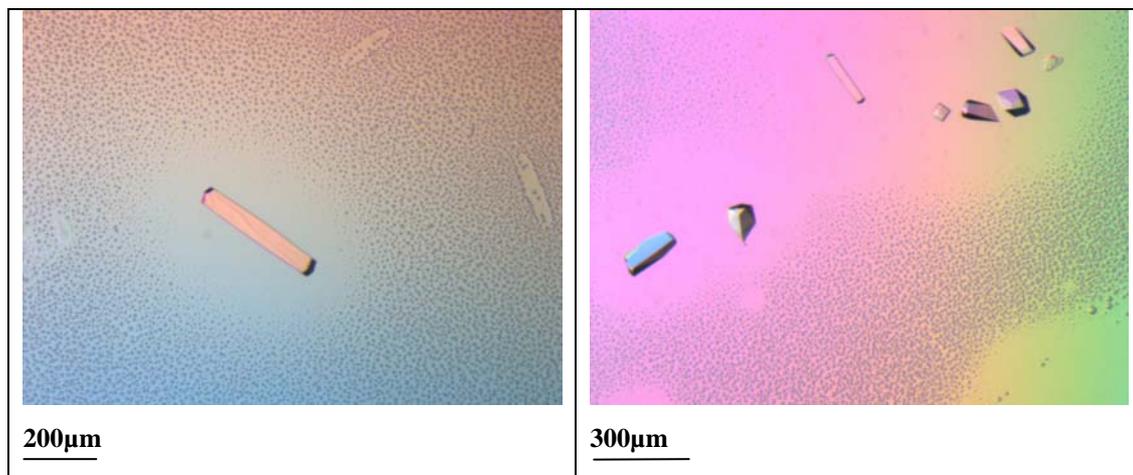


Figure 3: Tetragonal HEWL crystals grown in the presence of HEPES and MPD.

2 Data Collection

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

The relevant data collection parameters are given below:

wavelength	1.70 Å
detector distance:	70 mm
oscillation range/image:	1.0°
no of images:	180
exposure time/image:	7.4 sec
path to images:	exp5/data
image names:	exp5_lyso_ligands_1_###.img

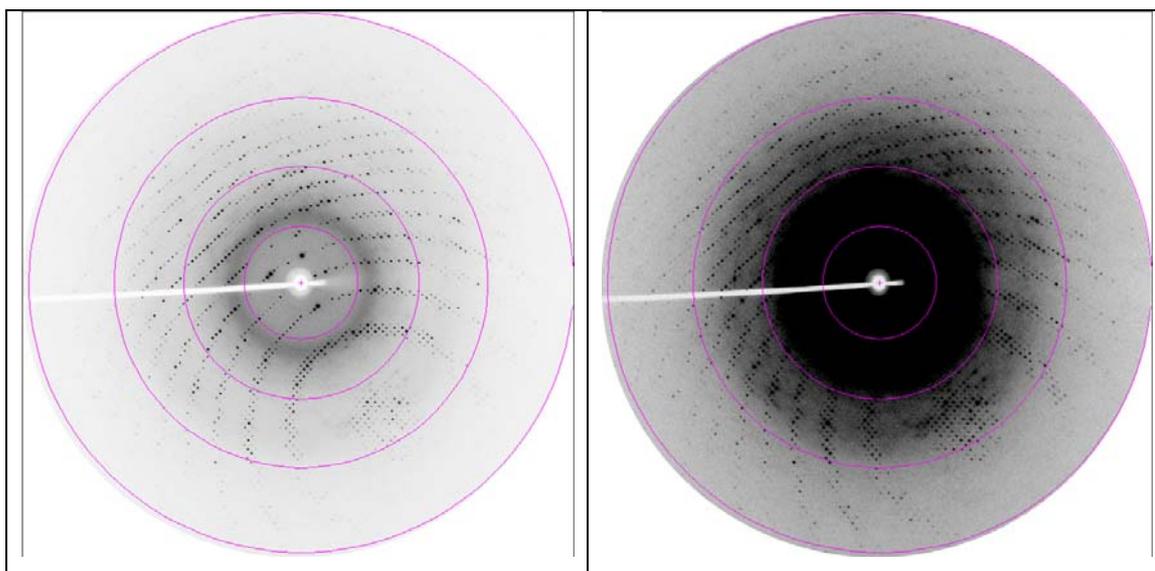


Figure 4: Diffraction image of tetragonal lysozyme displayed at different contrast levels. The resolution rings are at 7.0, 3.5, 2.3 and 1.8 Å, respectively.

3 Data Processing

The collected diffraction data were indexed, integrated and scaled using the program XDS (Kabsch, 1993, 2010a,b). XDS is simply run by the command *xds*. If a multi-processor machine is available, the command *xds_par* can be used, which calls a parallel version of XDS and consequently runs much faster. XDS needs only one input file, which must be called XDS.INP. No other name is recognized by the program. The file XDS.INP contains all relevant information about the data collection, from beam parameters to detector parameters and crystal parameters (if known) as well as the data collection geometry. In XDS.INP one can also define the steps through which the program should go. This is done by using the parameter JOBS. The following command, which is equivalent to JOBS= ALL would make XDS run through all eight steps XYCORR, INIT, COLSPOT, IDXREF, DEFPIX, XPLAN, INTEGRATE and CORRECT.

JOBS= XYCORR INIT COLSPOT IDXREF DEFPIX XPLAN INTEGRATE CORRECT

In the XYCORR step, tables of spatial correction factors are set up (if required). INIT calculates the gain of the detector and produces an initial background table. COLSPOT identifies strong reflections which are used for indexing. IDXREF performs the actual indexing of the crystal. DEFPIX identifies the regions on the detector surface which are used for measuring intensities, XPLAN helps to devise a data collection strategy, INTEGRATE integrates the reflection intensities of the whole data set and CORRECT scales and merges symmetry-related reflections and multiple measurements. It also prints out data processing statistics. After completing each individual step, a log-file with a name corresponding to the step (*STEP-name.LP*) is written.

Action 1: edit the supplied file XDS.INP and insert the relevant information about the data collection, namely the data collection wavelength, crystal-to-detector distance, the direct beam coordinates, the total number of images, and rotation increment per image and of course very importantly the path to and the names of the image files. XDS is able to recognize compressed images; therefore it is not necessary to unzip the data before using XDS. 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/*. If the space group and cell dimensions are known, the relevant information should be written into XDS.INP, if they are not known just set the parameter **SPACE GROUP NUMBER= 0**.

Action 2: run XDS until the indexing step, with the parameter JOBS set to:

JOBS= XYCORR INIT COLSPOT IDXREF

The output file `IDXREF.LP` contains the results of the indexing. It needs to be checked carefully whether the indexing is correct, since all subsequent steps assume the correctness of the indexing step. The most relevant parameters to look for are the STANDARD DEVIATION OF SPOT POSITION and the STANDARD DEVIATION OF SPINDLE POSITION. The first one should be in the order of 1 pixel, whereas the second one depends to some extent on the rotation increment per image but also on the mosaicity of the crystal. If it is 0.1° it is very good, if it is 0.5° it might still be ok, if it is larger than 1.0° the indexing has probably not worked. The table with the entries SUBTREE and POPULATION is also very interesting to look at. The first SUBTREE should have by a large margin more entries than all others. Also, the input parameters, such as the crystal-to-detector distance should after refinement not deviate too much from the input values.

The most common problem with the IDXREF step is that it often finishes with the message **!!! ERROR !!! INSUFFICIENT PERCENTAGE (<70%) OF INDEXED REFLECTIONS.** This means that less than 70% of the reflections that were collected in the COLSPOT step are not indexed, which may happen because of ice rings on the frames, split reflections or simply wrong input parameters. However, if all indicators of correct indexing are fine (see above) and no obvious errors can be identified then this message can be safely ignored and data processing can be continued.

If IDXREF was run with `SPACE_GROUP_NUMBER= 0`, an assumption of the correct Bravais lattice may be made at this stage. As a rule of thumb, choose the lattice of the highest possible symmetry, with a QUALITY OF FIT-value as low as possible (usually < 10). These numbers are printed in `IDXREF.LP` in the paragraph DETERMINATION OF LATTICE CHARACTER AND BRAVAIS LATTICE. Then, re-run the IDXREF step with the parameter `SPACE_GROUP_NUMBER` corresponding to the assumed Bravais lattice in `XDS.INP`. Alternatively, it is possible to make no assumption of the Bravais lattice at this point, and to simply continue with data integration. In this case the program will assume that space group determination should not be based on the metric symmetry of the lattice, but should be postponed to the CORRECT step (see below).

Nevertheless, if the true lattice is known, it should yield a good, i.e. low QUALITY OF FIT-number. For tetragonal HEWL, the correct space group is $P4_32_12$ (space group number 96) with unit cell parameters of $a=78.7$ and $c=37.1 \text{ \AA}$.

Action 3: After the determination of the Bravais lattice and the cell parameters all images need to be integrated and corrections (radiation damage, absorption, detector etc.) will have to be calculated. This can be done in a further XDS run.

JOBS= DEFPIX XPLAN INTEGRATE CORRECT

The CORRECT step produces a file called CORRECT.LP, which contains the statistics for the complete data set after integration and corrections. For the statistics to be meaningful, the correct Laue symmetry has to be established first. To this end, the CORRECT step compares the statistics in all possible Laue groups. The correct Laue group is the one with the highest symmetry, which at the same time still exhibits an acceptable $R_{r.i.m.}/R_{meas}$. CORRECT writes a file named XDS_ASCII.HKL, which contains the integrated and scaled reflections.

The CORRECT step also performs a refinement of all geometric parameters and the cell dimensions based on all reflections of the data set. These parameters may be more accurate than the ones obtained from the indexing step. Therefore, one may try to use the refined parameters and to re-run the last XDS job. In order to not overwrite the original results, it is advisable to save all current files to a temporary directory. Then, the file GXPARM.XDS should be renamed or copied to XPARAM.XDS and XDS be re-run. In case the original results are better, they can be copied back to the original directory.

While XDS will usually identify the correct Laue group, it does not determine the actual space group of the crystal. The decision about the existence of screw axes is left to the user. Indications, which screw axes may be present can be obtained from the table REFLECTIONS OF TYPE H,0,0 0,K,0 0,0,L OR EXPECTED TO BE ABSENT (*) in the file CORRECT.LP. Alternatively, the program POINTLESS (Evans, 2005) offers an automatic way of assigning the space group. POINTLESS can be run with the command *pointless XDSIN XDS_ASCII.HKL*. In the output the possible space groups together with their probabilities are given. Some space group ambiguity still remains at this stage, since it is impossible to distinguish between enantiomorphic space groups e.g. $P3_1$ and $P3_2$, or $P4_12_12$ and $P4_32_12$ just based on intensities. This ambiguity has to be resolved later during structure solution. The parameter SPACE_GROUP_NUMBER corresponding to the determined space group as well as the cell parameters should be entered into the file XDS.INP for running the next step.

Action 4: Finally, outlier reflections are identified by CORRECT by comparing their intensity to the average intensity in their respective resolution shells. These outliers may be removed, if there is a clear indication and reason for their existence: for example, ice rings often produce very strong reflections at specific d-spacings. The outliers are flagged as 'alien' in the file CORRECT.LP and their removal can simply be achieved by writing the outliers into a file called REMOVE.HKL. By re-running XDS with the command

JOBS= CORRECT

in XDS.INP, these outliers are then disregarded. This last action can be repeated until no more additional outliers are identified. However, the outlier removal has to be handled very carefully because strong reflections may also arise from non-crystallographic symmetry and in particular from the presence of pseudo-translation. A command to identify only the most extreme outliers would be `awk 'alien/ { if (strtonum($5) > 19) print $0 }' CORRECT.LP >> REMOVE.HKL`. This command will remove outliers only when their Z-score is above 19. Hints to suitable criteria for outlier rejection can be found in the XDSwiki (<http://strucbio.biologie.uni-konstanz.de/xdswiki>), where this question is treated specifically in the article “Optimization”.

Action 5:, The CORRECT step can be followed up by running the scaling program XSCALE, which is part of the XDS program package. This serves three purposes: a) the user may specify the limits of the resolution shells for which statistics should be printed, b) several XDS_ASCII.HKL files may be scaled together and c) correction factors for radiation damage may be applied to the data (see also the article “XSCALE” in the XDSwiki). XSCALE is run by simply typing `xscale` (or `xscale_par` to speed up the computation on a multi-processor machine) provided that a file XSCALE.INP defining the input and output files is present. As above in the CORRECT step, outliers may be rejected. 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.

Both CORRECT and XSCALE will produce all necessary output for assembling a table with all relevant data processing statistics, which is necessary for a publication.

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 ₃ 2 ₁ 2
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 [Å²]	31.4

Action 6: finally, the processed intensity file needs to be converted to certain file formats, which are used by other programs to perform the necessary structure determination steps. This can be achieved using the program XDSCONV, which can simply be run by using the command *xdscnv* provided that a file called XDSCONV.INP is present. XDSCONV.INP just needs to contain information about the name of the input file and about the name and type of the output file. If a CCP4-type file is required, XDSCONV reformats the reflection output file from XSCALE and creates an input file F2MTZ.INP for the final conversion of the reflection file to binary mtz-format, which is the standard format for all CCP4 programs (CCP4, 1994).

```
OUTPUT_FILE=lyso_ligands.hkl CCP4
```

```
INPUT_FILE=lyso_ligands.ahkl
```

To run the CCP4 programs F2MTZ and CAD, just type the two commands

```
f2mtz HKLOUT temp.mtz < F2MTZ.INP
```

and

```
cad HKLIN1 temp.mtz HKLOUT lyso_ligands_ccp4.mtz << eof
```

```
LABIN FILE 1 ALL
```

```
END
```

```
eof
```

Some CCP4 programs need the intensities of the Bijvoet pairs as input. For those the second parameter on the OUTPUT_FILE= line should be **CCP4_I** instead of **CCP4**. Alternatively, the file XDS_ASCII.HKL can be converted to mtz-format using the CCP4-programs COMBAT or POINTLESS (Evans, 2005) and this mtz-file can be used as an input file for the scaling program SCALA (Evans, 2005) in CCP4. More information on this can be found in the articles “Pointless” and “Scaling with SCALA” in the XDSwiki.

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; 2009) 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/Auto-Rickshaw/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 minimization step, a B-factor refinement and a second minimization 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 and identity 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.

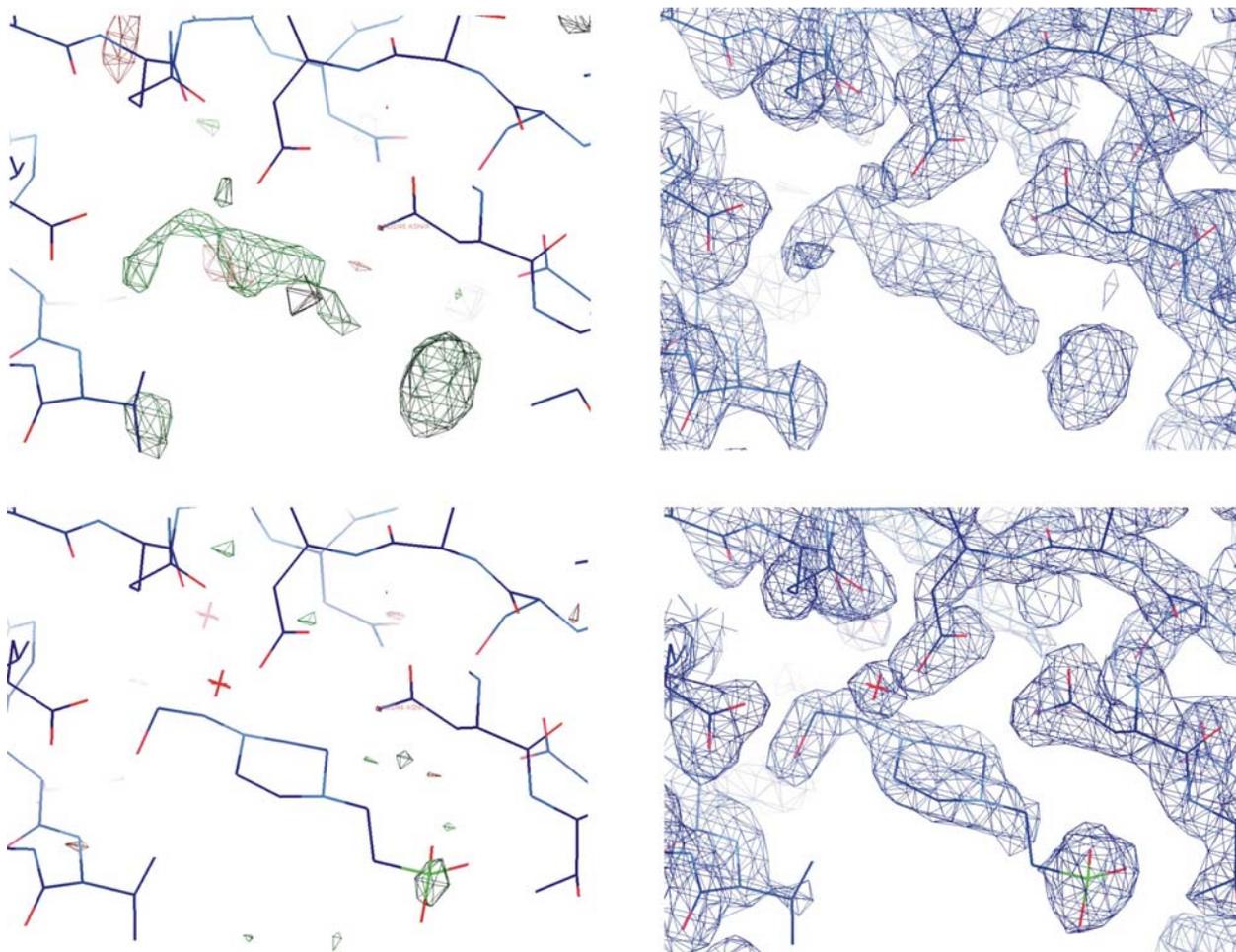


Figure 5: HEPES binding to HEWL. Top left panel: $(F_{\text{obs}} - F_{\text{calc}}, \alpha_{\text{calc}})$ -electron density contoured at $+2.0 \sigma$ (green) and -2.0σ (red), top right panel $(2F_{\text{obs}} - F_{\text{calc}}, \alpha_{\text{calc}})$ -electron density contoured at 1.0σ . In both cases the HEPES model was not included into the phase calculation. Bottom left panel: Anomalous difference Fourier electron density map contoured at $+3.0 \sigma$ 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 $(2F_{\text{obs}} - F_{\text{calc}}, \alpha_{\text{calc}})$ -electron density contoured at 1.5σ 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.
- Evans, P. (2005). *Acta Cryst.* **D62**, 72-82.
- Kabsch, W. (1993). *J. Appl. Cryst.*, **26**, 795-800.
- Kabsch, W. (2010a). *Acta Cryst.* **D66**, 125-132.
- Kabsch, W. (2010b). *Acta Cryst.* **D66**, 133-144.
- 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.
- Panjikar, S., Parthasarathy, V., Lamzin, V. S., Weiss, M. S. & Tucker, P. A. (2009). *Acta Cryst.* **D65**, 1089-1097.
- 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.