

A tutorial for learning and teaching macromolecular crystallography

Annette Faust,^a Santosh Panjekar,^a Uwe Mueller,^b Venkataraman Parthasarathy,^a Andrea Schmidt,^a Victor S. Lamzin^a and Manfred S. Weiss^{a*}

^aEMBL Hamburg Outstation, c/o DESY, Notkestrasse 85, D-22603 Hamburg, Germany, and ^bBESSY GmbH, BESSY-MX Group, Albert-Einstein-Strasse 15, D-12489 Berlin, Germany. Correspondence e-mail: msweiss@embl-hamburg.de

Five experiments have been designed to be used for teaching macromolecular crystallography. The three proteins used in this tutorial are all commercially available; they can be easily and reproducibly crystallized and mounted for diffraction data collection. For each of the five experiments the raw images and the processed data of a sample diffraction data set as well as the refined coordinates and phases are provided for teaching the steps of data processing and structure determination.

1. Introduction

Since the beginning of the 1990s, structural biology and macromolecular crystallography in particular have advanced enormously, as can be seen from the growth statistics of the Protein Data Bank (PDB; Berman *et al.*, 2000) shown in Fig. 1. While in the late 1980s one new macromolecular structure was determined every week worldwide, the rate has increased to about 20 new structures per day deposited with the PDB during the year 2007. In April 2008, the total content of the PDB exceeded 50 000 entries, the majority of which (> 85%) were determined by X-ray crystallographic techniques. This massive growth goes along with an increased importance of macromolecular crystallography in basic biological and medicinal research, but is hardly reflected adequately in the relevant biochemistry curricula at most universities around the world. Frequently, structural biology as a subject, or in more general terms the concept of thinking about biological problems in three dimensions, is introduced to students of biology or biochemistry at a relatively late stage in their career so that it is often too late for students to develop a deeper interest in this area. This has been realized and, consequently, there are voices advocating a much earlier introduction to the subject (Jaskólski, 2001). One of the problems, it seems, is that macromolecular crystallography taught as a lecture course tends to be a rather dry subject for students with a pure biology or biochemistry background. Since not all universities have a local macromolecular crystallography research group, access to experimental crystallography courses may be difficult or problematic. In order to address this issue, we have designed and assembled five macromolecular crystallography experiments, which can be carried out with relatively little effort. They are based on three commercially available protein samples, which can be easily and reproducibly crystallized and mounted for diffraction data collection. Should access to a data collection facility be not practical or difficult, this step can

be skipped and example data sets, which are provided for each project, can be used instead. These can then be integrated and scaled using the available software packages. The resulting diffracted intensities and anomalous differences, which are also provided as part of the tutorial, can then be used for structure determination according to standard crystallographic practice.

2. The five projects

The five experiments designed for this tutorial are (i) a sulfur single-wavelength anomalous diffraction (S-SAD) structure determination of cubic insulin, (ii) a bromide multiple-wavelength anomalous diffraction (Br-MAD) structure determination of thaumatin, (iii) a structure determination by molecular replacement using monoclinic lysozyme, (iv) the identification of bound solvent ions in lysozyme using a

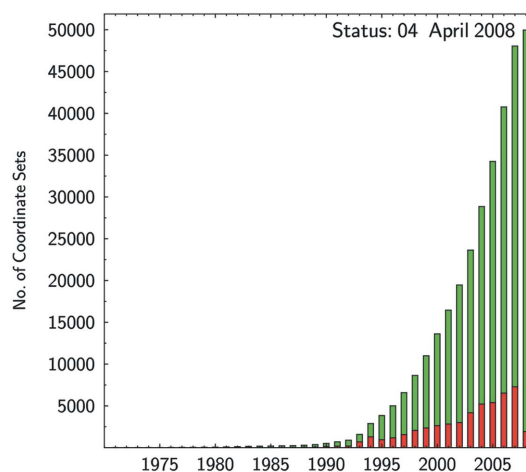


Figure 1
 Growth and content of the PDB (Berman *et al.*, 2000) as of April 2008. The red bars denote the annual deposits and the green bars the total content of the PDB.

longer-wavelength data set and (v) the identification of a weakly bound ligand at the active site of lysozyme. These five cases (Fig. 2) represent some of the most common scenarios encountered in structural biology using macromolecular crystallographic techniques.

2.1. S-SAD of cubic insulin

Insulin regulates the cellular uptake, utilization and storage of glucose, amino acids and fatty acids and inhibits the breakdown of glycogen, protein and fat. It is a two-chain polypeptide hormone produced by the β cells of pancreatic

islets (Voet *et al.*, 2006). The two chains comprise a total of 51 amino acids. Three disulfide bonds hold the two chains together: one intra-chain SS bridge between Cys6 and Cys11 in chain A, and two inter-chain SS bridges, one between Cys7 from chain A and Cys7 from chain B, and the other between Cys20 from chain A and Cys19 from chain B (Fig. 2a).

Experimental phase determination using SAD from the S atoms inherently present in nearly all protein molecules has in the past few years experienced a huge boost in popularity. After the initial success with the small protein crambin (46 amino acids, six S atoms) by Hendrickson & Teeter (1981), it took 18 years until the method was revived by Dauter *et al.* (1999), who were able to demonstrate by the successful structure determination of hen egg-white lysozyme (HEWL) based on the anomalous scattering of the protein S atoms and surface-bound chloride ions alone that the requirements in terms of data quality are less strict than was commonly believed and that they are indeed manageable. Since then, approximately 100 structures have been obtained using this so-called sulfur-SAD or S-SAD approach. In view of the fact that the anomalous signal from the light atoms at the typically used wavelengths in macromolecular crystallography is small, it has been suggested and experimentally verified that diffraction data collection at somewhat longer wavelengths may be beneficial (Djinović Carugo *et al.*, 2005; Mueller-Dieckmann *et al.*, 2005; Weiss *et al.*, 2001). However, while a larger anomalous signal may be obtained at longer X-ray wavelengths, additional experimental complications arising mainly from X-ray absorption have to be dealt with. In this experiment, cubic crystals of bovine insulin are used for experimental phase determination using diffraction data collected at a wavelength of $\lambda = 1.77 \text{ \AA}$.

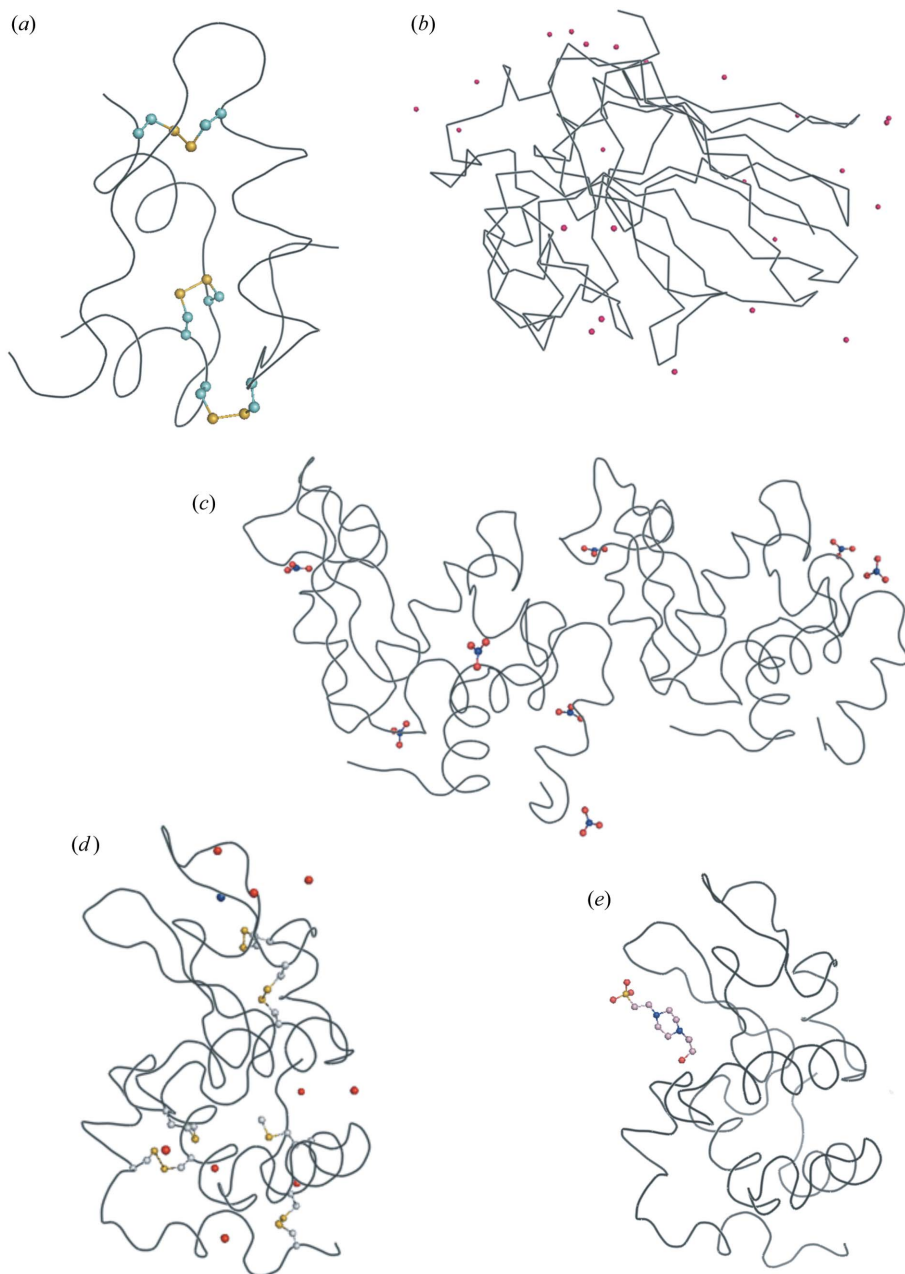


Figure 2
The five projects. (a) S-SAD on cubic insulin, (b) Br-MAD on thaumatin, (c) molecular replacement on monoclinic lysozyme, (d) solvent ion identification in tetragonal lysozyme and (e) ligand identification in the MPD form of tetragonal lysozyme.

2.2. Br-MAD of thaumatin

Thaumatococcus daniiellii is a mixture of three intensely sweet proteins isolated from the seed vessel of the Katemfe plant (*Thaumatococcus daniiellii*). It is about 1000 times sweeter than sucrose on a weight basis, 100 000 times on a molar basis, and is therefore used in the food industry as a sweetener. The commercially available thaumatin is a mixture of thaumatin I and thaumatin II with

traces of other sweet proteins. The amino acid sequence of thaumatin contains 207 residues, where thaumatin I and II differ from one another in five amino acids only. As it is a mixture, it is hard to examine the ratio between thaumatin I and II in the crystal structure during refinement. All PDB entries for thaumatin (Fig. 2*b*) are therefore modelled using the thaumatin I sequence (Ko *et al.*, 1994).

Over the past two decades, MAD has been the standard method for *de novo* structure determination in macromolecular crystallography (Hendrickson & Ogata, 1997; Hendrickson, 1999). In MAD, the wavelength-dependent anomalous scattering properties from heavy atoms that are part of or bound to the macromolecule of interest are utilized. The heavy atoms can be directly incorporated into the protein (*e.g.* seleno-methionine derivatives or metal-containing proteins) or they can be soaked into the crystal. MAD experiments are carried out at different X-ray energies around an absorption edge of the heavy atom where the anomalous scattering factors of the heavy atom are significantly different from each other. Up to four diffraction data sets are collected, at the peak wavelength where $\Delta f''$ reaches its maximum, at the inflection point wavelength where $\Delta f'$ reaches its minimum, and away from the absorption edge at wavelengths at least 100 eV remote from the absorption edge on the high-energy and low-energy sides.

2.3. Molecular replacement of lysozyme

Lysozyme is a 129 amino acid enzyme that dissolves bacterial cell walls by catalysing 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). In addition, large amounts of lysozyme can be found in egg whites. The crystal structure of 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 crystal forms (for an overview, see Brinkmann *et al.*, 2006) and has become a standard substance for methods development and teaching purposes.

In this experiment, the structure of monoclinic HEWL (Fig. 2*c*) 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).

2.4. Identification of bound solvent ions in lysozyme

Recently, it has been observed, by examining longer-wavelength diffraction data sets collected from 23 different

protein crystals, that most macromolecular crystal structures contain some sort of solvent ions (Mueller-Dieckmann *et al.*, 2007) more or less tightly bound to their surface. This is in stark contrast to the findings in the PDB (Berman *et al.*, 2000), where in most cases only water molecules were modelled to be bound to the surface of the crystallized macromolecules. This suggests that in many structures surface-bound ions might have mistakenly been assigned as water molecules. One way to remedy the situation is to complement every structure determination with a longer-wavelength data set, in which the presence of peaks in the anomalous difference Fourier electron density map is clearly able to distinguish water molecules from ions such as sulfate, phosphate, chloride or potassium. In this experiment, good quality anomalous differences are collected from a tetragonal crystal of HEWL at a wavelength close to the copper $K\alpha$ -edge ($\lambda = 1.5418 \text{ \AA}$), in order to identify the complete anomalously scattering substructure of the protein (Fig. 2*d*).

2.5. Identification of a weakly bound ligand at the active site of lysozyme

Knowledge of the three-dimensional structure of a drug target protein is the first step in crystallographic structure-based drug design. The architecture of the active site of the target molecule can serve as the basis for the design of potential molecules, which can bind to the active site and may inhibit the enzymatic reaction of the target enzyme. 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 purpose, advantage can be taken of the recent advances towards automatization and high throughput in crystallization and data collection. Data collected from crystals of a particular protein soaked with many different ligands can be quickly evaluated using the direct refinement approach, *i.e.* by skipping the molecular replacement step. This method is also more generally applicable and allows the identification of inhibitors, substrate or products bound to a macromolecule in order to evaluate reaction mechanisms and other things. In this experiment, tetragonal crystals of HEWL grown from HEPES buffer in the presence of 2-methyl-2,4-pentanediol (MPD) are examined in order to identify a weakly bound HEPES molecule at the active site of the enzyme (Fig. 2*e*).

3. Crystallization

All crystallization experiments were carried out using the hanging-drop method at room temperature (293 K) in the screw-cap crystallization plates (EasyXtal Tool) available from Nextal (now Qiagen).

3.1. Insulin

Bovine insulin ($M_r = 5.73 \text{ kDa}$, Sigma catalogue No. I5500) crystals were prepared by mixing 4 μl of protein solution (20 mg ml^{-1} insulin in 20 $\text{mM Na}_2\text{HPO}_4$ and 10 mM

Na₃EDTA pH 10.0–10.6) and 4 µl of reservoir solution containing 225–350 mM Na₂HPO₄/Na₃PO₄ (pH 10.0–10.6) and 10 mM Na₃EDTA and equilibrating the drop against the reservoir solution. Crystals belonging to the cubic space group *I*2₁3 with the unit-cell parameter $a = 78.0 \text{ \AA}$ grew within a few days up to a final size of 100–300 µm (Fig. 3*a*). All chemicals used for the experiment were from Sigma, unless specified otherwise.

3.2. Thaumatin

Thaumatin ($M_r = 22.2 \text{ kDa}$, Sigma–Aldrich catalogue No. T-7638) crystals (Fig. 3*b*) grew within a few days after mixing 2 µl of protein solution [15 mg ml⁻¹ in 0.1 M *N*-(2-acetamido)-iminodiacetic acid (ADA) pH 6.5] and 2 µl of reservoir solution (0.1 M ADA pH 6.5, 1 M sodium/potassium tartrate) and equilibrating the drop against the reservoir. The tetragonal crystals (space group *P*4₁2₁2) exhibit unit-cell parameters $a = 57.7$, $c = 150.2 \text{ \AA}$ and diffract X-rays to beyond 1.5 Å resolution (Mueller-Dieckmann *et al.*, 2005). All chemicals used for the experiment were from Sigma, unless specified otherwise.

3.3. Monoclinic lysozyme

Monoclinic HEWL ($M_r = 14.6 \text{ kDa}$, Fluka catalogue No. 62970) crystals (Fig. 3*c*) were prepared according to a recipe described by Saraswathi *et al.* (2002) by mixing 12 µl of protein solution (20 mg ml⁻¹ 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% (w/v) sodium nitrate and equilibrating the drop against the reservoir. Crystals belonging to the monoclinic space group *P*2₁ with unit-cell parameters $a = 27.5$, $b = 62.4$, $c = 59.6 \text{ \AA}$, $\beta = 90.5^\circ$ grew within a few days. All chemicals used for the experiment were from Sigma, unless specified otherwise.

3.4. Tetragonal lysozyme (NaCl form)

Tetragonal crystals of HEWL were grown as described by Weiss *et al.* (2000) by mixing 4 µl of protein solution (30 mg ml⁻¹ 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. Crystals belonging to space group *P*4₃2₁2 and exhibiting the usual unit-cell parameters of $a = 78.8$, $c = 37.2 \text{ \AA}$ appeared within a few days (Fig. 3*d*). All chemicals used for the experiment were from Sigma, unless specified otherwise.

3.5. Tetragonal lysozyme (HEPES/MPD form)

Another tetragonal crystal form of HEWL was grown by mixing 4 µl of protein solution (30 mg ml⁻¹ 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 *P*4₃2₁2) appeared within a few days and exhibited the unit-cell parameters $a = 78.7$, $c = 37.1 \text{ \AA}$. All chemicals used for the experiment were from Sigma, unless specified otherwise.

4. Diffraction data collection

Prior to the diffraction experiment the crystals were mounted in nylon loops, cryo-protected if necessary and shock-cooled in a cold nitrogen gas stream directly on the beamline. Insulin crystals were cryo-protected in a solution containing 250 mM Na₂HPO₄/Na₃PO₄ pH 10.2, 10 mM Na₃EDTA and 30% (v/v) glycerol (Sigma–Aldrich). Thaumatin crystals were simultaneously derivatized and cryo-protected by soaking them for a few seconds in a solution containing 1.0 M NaBr and 25% (v/v) glycerol. Monoclinic and tetragonal HEWL crystals (NaCl form) were cryo-protected using dry paraffin oil (Fluka catalogue No. 76235), and tetragonal HEWL crystals (MPD form) were used directly without further cryo-protection. The diffraction data from the five crystals were collected on the beamlines X12 (EMBL Hamburg, c/o DESY, Hamburg, Germany), BL14.1 and BL14.2 (Freie Universität Berlin and BESSY, Berlin-Adlershof, Germany). Beamlines X12 and BL14.1 are equipped with a MAR-Mosaic CCD detector (225 mm) mounted on a MARdtb goniometer system, and beamline BL14.2 is equipped with a smaller MARCCD (165 mm), which is also mounted on a MARdtb.

The wavelengths for data collection were chosen on the basis of the requirements of the project. For the S-SAD experiment on insulin, data were collected at a slightly longer wavelength of $\lambda = 1.77 \text{ \AA}$ in order to enhance the anomalous signal of the S atoms for the subsequent structure determination steps. The same

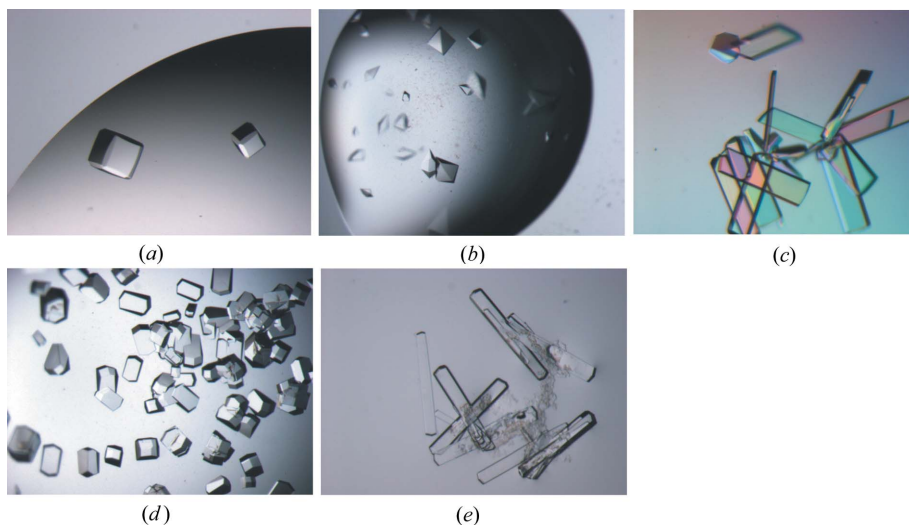


Figure 3

Crystals of (a) bovine insulin, (b) thaumatin, (c) monoclinic lysozyme, (d) tetragonal lysozyme NaCl form and (e) tetragonal lysozyme MPD form.

Table 1

Data collection parameters for the five projects.

Project	Insulin (Exp. 1)	Thaumatin (Exp. 2)		High-energy remote	Low-energy remote	Lysozyme (Exp. 3)	Lysozyme (Exp. 4)	Lysozyme (Exp. 5)
		Peak	Inflection point					
No. of crystals	1			1		1	1	1
Beamline	BL14.1			X12		BL14.2	BL14.1	BL14.1
Wavelength (Å)	1.770	0.9188	0.9196	0.9134	0.9252	0.900	1.5418	1.700
Crystal–detector distance (mm)	50			200		100	100	70
Rotation range per image (°)	1.0			0.5		1.0	1.0	1.0
Total rotation range (°)	180			180		360	360	180
Exposure time per image (s)	2.6	4.5	4.1	4.3	4.3	3.4	3.4	7.4

rationale was applied to the experiment on the identification of a weakly bound ligand in tetragonal lysozyme ($\lambda = 1.70 \text{ \AA}$). The weak but significant anomalous signal from the S atom of HEPES helps to identify the bound ligand with confidence. The four wavelengths at which the thaumatin MAD data were collected were determined on the basis of an X-ray fluorescence scan carried out at the bromine *K*-absorption edge. The scan (see example picture in the accompanying tutorial) can be evaluated manually or automatically using the program *CHOOCH* (Evans & Pettifer, 2001). The wavelengths chosen for data collection were those at the peak of the bromine absorption edge where $\Delta f''$ is maximal, at the inflection point where $\Delta f'$ is minimal, and away from the absorption edge at wavelengths on the high-energy and low-energy sides of the absorption edge. Even though most MAD data collections nowadays consist of three or fewer wavelengths, we decided to include all four wavelengths in the experiment. It is mostly the data set collected at the low-energy side of the edge that is omitted, because it provides close to no additional signal for phasing. Nevertheless, it is instructive to see that the low-energy remote data set is often the one of the highest quality, because the lack of X-ray fluorescence results in a much reduced background in the diffraction images. The wavelengths for the two remaining experiments were chosen arbitrarily. With $\lambda = 1.00 \text{ \AA}$ (for the molecular replacement experiment on monoclinic lysozyme) and $\lambda = 1.54 \text{ \AA}$ (for the experiment on the identification of bound solvent ions), they merely reflect a typical synchrotron and a typical home source diffraction experiment, respectively.

All relevant data collection parameters are summarized in Table 1.

5. Data processing

All data sets were indexed, integrated and scaled using the program *XDS* (Kabsch, 1993). For teaching purposes, the detailed protocol employed for data processing is described in the following, although much of this information can be gathered from the *XDS* program manual. *XDS* needs only one input file called *XDS.INP*. *XDS* is able to recognize compressed images, but in the file *XDS.INP* the image name given must not include the zipping-format extension. Furthermore, *XDS* has a very limited string length (80 characters) to describe the path to the images. Before running

XDS, the file *XDS.INP* has to be edited so that it contains the correct data collection parameters. To let the program automatically determine the space group and the cell parameters, the space-group number in *XDS.INP* should be set to 0 and the parameter *JOBS* to *XYCORR INIT COLSPOT IDXREF*. The results of the indexing procedure and the relevant parameters can then be found in the output file *IDXREF.LP*. For insulin, for instance, the correct space group is $I2_13$ (space-group number 199) with the unit-cell parameter $a = 78.0 \text{ \AA}$. After the determination of the space group and the cell parameters, all images are integrated in a second *XDS* run with the space-group number set to the correct number and the parameter *JOBS* to *DEFPIX XPLAN INTEGRATE CORRECT*. If the cell parameters and the space group are known before the first run one can run *XDS* with *JOBS* = *ALL*, which will make a second run obsolete. After integration, scaling and various corrections, the output file *CORRECT.LP* contains the statistics for the complete data set. A file named *XDS_ASCII.HKL* is written out, which contains all the integrated and scaled reflections. The program *XSCALE* carries out the final scaling and merging steps and writes out a *.ahkl file, which can be converted with *XDSCONV* for use within the *CCP4* suite (Collaborative Computational Project, Number 4, 1994) or other programs. *XDSCONV* creates an input file *F2MTZ.INP* for the final conversion to the binary mtz format used by *CCP4*. Alternatively, the *XDS_ASCII.HKL* file can be converted using the program *COMBAT* (Collaborative Computational Project, Number 4, 1994) and the resulting mtz file can be used as input file for the *CCP4* program *SCALA* (Collaborative Computational Project, Number 4, 1994) for scaling and merging. A detailed description of the data processing including the relevant lines of the input files is provided in the tutorials, which can be downloaded from http://www.embl-hamburg.de/Xray_Tutorial and <http://www.mx.bessy.de/bessy-ws/bessy.html>. The data processing statistics obtained with *XDS* for the various data sets are given in Table 2.

6. Structure determination

Because the focus of the workshop for which this tutorial was originally designed was on crystal mounting, data collection and processing, the actual structure determination was carried out automatically mainly to validate the success of the

Table 2

Data processing statistics for the five projects.

The numbers in parentheses define the outermost resolution shell used.

Project	Thaumatococcus (Exp. 2)					Lysozyme (Exp. 3)	Lysozyme (Exp. 4)	Lysozyme (Exp. 5)
	Insulin (Exp. 1)	Peak	Inflection point	High-energy remote	Low-energy remote			
Resolution range (Å)	50.0–1.60 (1.70–1.60)	50.0–1.82 (2.00–1.82)	50.0–1.82 (2.00–1.82)	50.0–1.82 (2.00–1.82)	50.0–1.83 (2.00–1.83)	50.0–1.35 (1.40–1.35)	50.0–1.80 (2.00–1.80)	50.0–2.00 (2.10–2.00)
Space group	$I2_13$		$P4_12_12$			$P2_1$	$P4_32_12$	$P4_32_12$
Unit-cell parameters (Å, °)	$a = 78.0$		$a = 57.80, c = 149.96$			$a = 27.40, b = 62.30,$ $c = 59.50, \beta = 90.50$	$a = 78.80, c = 37.17$	$a = 78.70, c = 37.10$
Mosaicity (°)	0.14	0.24	0.24	0.24	0.24	0.50	0.20	0.50
Total No. of reflections	208 859	331 063	330 552	335 218	324 973	317 114	286 190	105 988
Unique reflections	20 226	43 316	43 274	43 491	42 550	43 604	20 602	15 059
Completeness (%)	99.9 (100.0)	99.6 (98.3)	99.5 (97.9)	100.0 (100.0)	97.8 (91.2)	99.2 (93.3)	99.3 (97.6)	100.0 (100.0)
Redundancy	10.3 (8.3)	7.6 (7.1)	7.6 (7.3)	7.7 (7.6)	7.7 (7.3)	7.3 (5.5)	13.9 (11.6)	7.0 (6.6)
$I/\sigma(I)$	48.1 (13.6)	13.1 (2.9)	14.4 (3.3)	13.7 (2.9)	17.4 (4.7)	27.5 (3.9)	43.1 (18.1)	24.4 (3.0)
$R_{r.i.m.}/R_{meas}$ (%)†	3.7 (17.0)	15.2 (80.5)	13.3 (70.4)	14.7 (82.4)	10.1 (48.6)	4.1 (48.0)	4.6 (14.0)	7.0 (80.2)
Overall Wilson B factor (Å ²)	22.1	21.9	22.2	22.6	22.6	19.6	22.3	33.3

 † $R_{r.i.m.}/R_{meas} = 100 \sum_{hkl} [N/(N-1)]^{1/2} \sum_i |I_i(hkl) - \langle I(hkl) \rangle| / \sum_{hkl} \sum_i I_i(hkl)$, where $I_i(hkl)$ is the intensity of the observation i of the reflection hkl and N is the redundancy (Diederichs & Karplus, 1997; Weiss, 2001).

diffraction experiment. For this purpose, the respective protocols of the EMBL-Hamburg automated crystal structure determination platform Auto-Rickshaw (Panjikar *et al.*, 2005) were utilized. Auto-Rickshaw can be accessed from outside EMBL under <http://www.embl-hamburg.de/AutoRickshaw/> (a free of charge registration may be required; please follow the instructions on the web page). Of course, the tutorial may also be used with the emphasis placed on teaching the process of structure determination. In such cases, the subsequent steps of a structure determination can be performed the conventional manual way and the numbers provided here may simply serve as a rough guide.

For the first two experiments (S-SAD on insulin and Br-MAD on thaumatococcus) phases need to be derived from the measured anomalous differences. This is typically carried out as a sequence of steps. The first step is the calculation of ΔF values in the case of SAD or F_A values in the case of MAD. This step can be performed, for instance, with the program *SHELXC* (Sheldrick *et al.*, 2001; Sheldrick, 2008). *SHELXC* will also write an input file for the subsequent program *SHELXD* (Schneider & Sheldrick, 2002), which is often used to determine the anomalously scattering substructure. *SHELXD* uses a dual-space approach for that purpose. Since there is a random element in starting a job, it can be run many times (a typical number of cycles is 100, but it is often worthwhile to run up to 10 000 cycles). A correct solution can then be identified by looking at the two correlation coefficients CC(All) and CC(weak). Values of 30.0 and 15.0 probably indicate a correct solution, although it may be the case that correct solutions appear with lower values. The next step is the actual phase calculation and the improvement of the phase by density modification. This can be achieved by using the program *SHELXE* (Sheldrick, 2002), or programs such as *MLPHARE* and *DM* (Collaborative Computational Project, Number 4, 1994) or others. At this point, the correct hand of the substructure also has to be established. Once the phases have been determined an electron density map can be calcu-

lated, for instance inside the program *COOT* (Emsley & Cowtan, 2004), and displayed and examined for parts that can be interpreted in terms of an amino acid chain or a secondary structure element. With high enough resolution at hand, however, it is also possible to attempt a completely automatic density interpretation and model building using the program *ARP/wARP 7.0* (Perrakis *et al.*, 1999; Morris *et al.*, 2002).

In experiment 3 (MR on monoclinic lysozyme), structure solution will be performed by molecular replacement. In order to make the experiment realistic the search model should ideally be derived from a different crystal form of HEWL. A good model would, for instance, be the high-resolution structure of tetragonal HEWL (PDB code 193l; Vaney *et al.*, 1996). The model should be stripped of all water molecules and used as input to the program *MOLREP* (Collaborative Computational Project, Number 4, 1994; Vagin & Teplyakov, 1997) or any other molecular replacement program. Since two molecules are expected in the asymmetric unit, the program will try to orient and position the first molecule, fix this one, and then try to orient and position the second. Once this is achieved the model can be subjected to refinement using *CNS* (Brünger *et al.*, 1998) or *REFMAC5* (Murshudov *et al.*, 1997).

For experiments 4 and 5 (bound solvent ions and ligand identification in HEWL), the structure can be determined by just taking another HEWL structure of the same crystal form from the PDB (Berman *et al.*, 2000). Again, a good model would be the PDB entry 193l (Vaney *et al.*, 1996) or the PDB entry 1dpw (Weiss *et al.*, 2000). Since the packing of the molecules is the same and the cell dimensions are very similar, no molecular replacement needs to be performed and one can proceed directly to refinement. It is advisable, however, to first carry out rigid-body refinement at lower resolution, for instance 3.0 Å, and then the normal restrained refinement to the maximum resolution of the data collected. The phases derived from the refined models will then be used to calculate the $(2F_{obs} - F_{calc}, \alpha_{calc})$, the $(F_{obs} - F_{calc}, \alpha_{calc})$ and the $(\Delta F, \alpha_{calc} - 90^\circ)$ electron density maps, which can be displayed in *COOT*

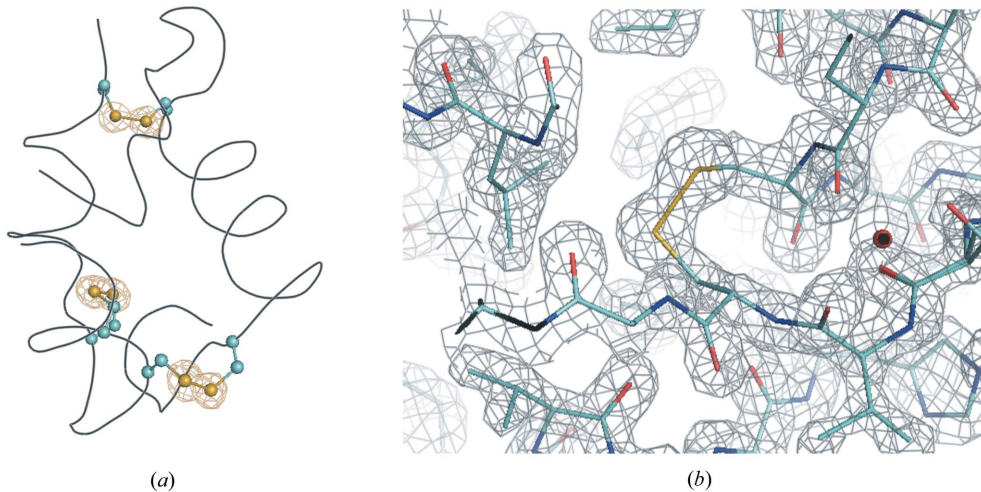


Figure 4

(a) Anomalous difference map of insulin superimposed on the final refined model. The map is contoured at 8.0σ . (b) Snapshot of the experimental electron density map of insulin superimposed onto the final refined model. The experimental map is contoured at 1.5σ .

in order to identify the complete anomalously scattering substructure (experiment 4) or the weakly bound ligand (experiment 5). Alternatively, one may try to use the anomalous differences to attempt experimental phasing for structure determination as described for experiment 1 above.

6.1. Insulin

The structure can be solved using the SAD protocol of Auto-Rickshaw run in the advanced version. The input diffraction data (file XDS_ASCII.HKL) were uploaded and then prepared and converted for use in Auto-Rickshaw using programs of the CCP4 suite. ΔF values were calculated using the program SHELXC. On the basis of an initial analysis of the data, the maximum resolution for substructure determination and initial phase calculation was set to 1.8 \AA . All of the six heavy atoms requested were found using the program SHELXD with correlation coefficients CC(All) and CC(weak) of 53.3 and 32.2, respectively, and with a clear drop

in occupancy after site No. 6. This indicates that the correct solution was most likely found. The correct hand for the substructure was determined using the programs ABS (Hao, 2004) and SHELXE. Initial phases were calculated after density modification using the program SHELXE and extended to 1.60 \AA resolution. 90% of the model was built using the program ARP/wARP 7.0. The complete Auto-Rickshaw run in the advanced version took around 20 min. The model was then further modified using COOT and refined using REFMAC5. Panels (a) and (b) of Fig. 4 show snapshots of the final model

superimposed with the anomalous difference electron density map and the experimental electron density map after density modification in SHELXE.

6.2. Thaumatin

The structure was solved using the 3W-MAD protocol of Auto-Rickshaw using the peak, inflection point and high-energy remote-wavelength data sets and the amino acid sequence of thaumatin. The input diffraction data (file XDS_ASCII.HKL) were uploaded and then prepared and converted using programs of the CCP4 suite. F_A values were calculated using the program SHELXC. On the basis of an initial analysis of the data, the maximum resolution for substructure determination and initial phase calculation was set to 2.4 \AA . All of the 20 heavy atoms requested were found using the program SHELXD with correlation coefficients CC(All) and CC(weak) of 30.9 and 13.5, respectively. The correct hand for the substructure was determined using the

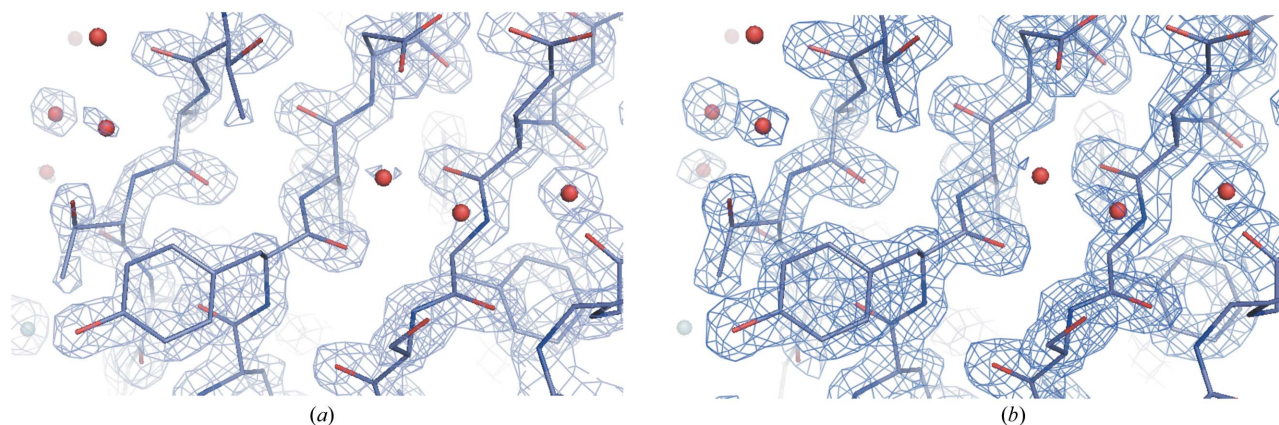


Figure 5

Structure solution of thaumatin. (a) The experimental map after density modification followed by an initial model building step in ARP/wARP 7.0, and (b) experimental map with the refined model. The maps are contoured at 2.0σ .

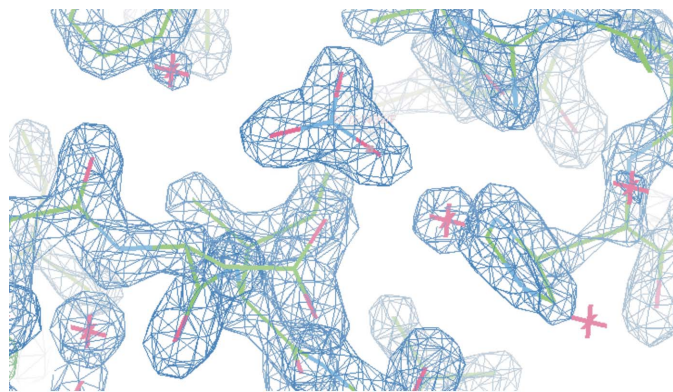


Figure 6

Molecular replacement on monoclinic lysozyme. The final model-phased electron density map has been superimposed onto the refined structure. Some nitrate ions bound to the surface of the protein are clearly visible. The ($2F_{\text{obs}} - F_{\text{calc}}$, α_{calc}) map (blue) is contoured at 1.5σ .

programs *ABS* and *SHELXE*. Initial phases were calculated after density modification using the program *SHELXE*. 99% of the model was built using the program *ARP/wARP*. The model was then further modified using *COOT* and refined using *REFMAC5*. Figs. 5(a) and 5(b) show the experimental map after density modification followed by a model building step in *ARP/wARP 7.0* and the model-phased electron density map after the final refinement.

6.3. Monoclinic lysozyme

The structure was solved using the MR protocol of Auto-Rickshaw with tetragonal HEWL (PDB code 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 performed using *MOLREP* with a resolution cut-off of 4.0 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* in four consecutive steps: rigid-body refinement, a minimization step, *B*-factor refinement and a second minimization step. At this point the *R* and R_{free} values were 24.9 and 33.5%, respectively. Further refinement was carried out 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*. Fig. 6 shows the final electron density with some nitrate ions clearly visible.

6.4. Tetragonal lysozyme (NaCl form)

The structure was solved using the MR protocol of Auto-Rickshaw with tetragonal HEWL (PDB code 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. Because the cell parameters of the model structure and the target structure were very similar, Auto-Rickshaw skipped the *MOLREP* step

Table 3

List of the peaks in the anomalous difference Fourier electron density map of tetragonal lysozyme (NaCl form) above 4.5σ and their assignment to atoms and ions of the structure.

Peak No.	Peak height/r.m.s.	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	10.1	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	5.0	Cl-9

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 minimization step, *B*-factor refinement and a second minimization step. At this point the *R* and R_{free} values were 22.6 and 32.4%, respectively. Further refinement was then carried out 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 Cl and Na atoms. Table 3 contains the highest peaks found in the anomalous difference Fourier electron density map (threshold 4.5σ above the mean value of the map). Nine chloride ions can be identified, in addition to the intrinsically present ten S atoms of HEWL. Fig. 7 shows the refined structure of HEWL overlaid with the anomalous map indicating the bound chloride ions.

6.5. Tetragonal lysozyme (HEPES/MPD form)

The structure was solved using the MR protocol of Auto-Rickshaw with tetragonal HEWL (PDB code 1dpw; Weiss *et al.*, 2000) as a starting model. The input diffraction data (XDS_ASCII.HKL) were uploaded and then prepared and converted using programs of the *CCP4* suite. Because the cell parameters of the model and the target structure were very similar, Auto-Rickshaw skipped the *MOLREP* step and proceeded directly to refinement. The refinement was performed to a resolution cut-off of 3.0 Å in *CNS* in four consecutive steps: rigid-body refinement, a minimization step, *B*-factor refinement and a second minimization step. At this point the *R* and R_{free} values were 22.2 and 32.8%, respectively. Further refinement was carried out in *REFMAC5* using all available data to *R* and R_{free} values of 27.4 and 31.5%. The model was completed using *COOT* and further refined using *REFMAC5*. Residual electron density was found for a bound



Figure 7

The refined structure of the NaCl form of tetragonal HEWL as a wire model superimposed onto the anomalous difference Fourier electron density map (red for the nine Cl atoms, dark blue for Na, yellow for S) contoured at 4.0σ , which clearly identifies the bound chloride ions.

HEPES molecule (Fig. 8). To verify the presence and the position of the ligand, an anomalous difference Fourier electron density map was calculated to identify the S atom (Fig. 8). Fig. 8 also shows the final electron density of HEPES after refinement.

7. Challenges and didactic value

The five experiments have been designed such that they can easily be carried out by beginners in the field but they are also challenging in some respect. The three proteins used for crystallization are all available commercially, and the crystallization experiments themselves usually work reliably and reproducibly.

The diffraction data collection experiments can be carried out on any X-ray source, be it a synchrotron beamline or a home source. Except for the bromide-MAD experiment, all experiments should work just as well with data collected from a home source at the Cu $K\alpha$ wavelength. As a matter of fact, the wavelength for the experiment of identification of ions bound to the surface of HEWL (experiment 4) has been chosen so as to mimic data collection at a typical home source. The resolution for data collection is not very critical. At a synchrotron beamline, all crystals typically diffract much further than 1.8 Å. It is advisable, however, to choose a maximum resolution of at least 2.0 Å for the diffraction

experiment, since it makes the following structure determination steps easier. The diffraction data provided with the tutorial were collected as part of a workshop on *Diffraction Data Collection Using Synchrotron Radiation* (<http://www.embl-hamburg.de/workshops/2007/diffraction/>). The crystals were mounted by the students of the workshop and the data were collected by them, albeit under the guidance of and supervision by the tutors. The intention was not to collect perfect data but real case data sets, which do pose some challenges for data processing and structure determination.

Data processing is mostly straightforward for all the presented cases. The only somewhat challenging project in this respect is the monoclinic HEWL data set, where the β angle is close to 90° and the b and c axes are close in length, so that mis-indexing is easily possible. During the workshop, the only data processing program used was *XDS*, but any other popular data processing program suite will work as well. It might even be an interesting teaching experiment to have different groups of students use different programs for processing the same data set and have them compare the results, both in terms of the resulting merging statistics and in terms of whether the structure can be determined or not. As a rough guide to what kind of merging statistics can be expected, those obtained from *XDS* are provided in Table 2.

Structure determination has only been attempted automatically using the Auto-Rickshaw platform (Panjekar *et al.*, 2005), because the focus of the workshop was on crystal mounting, data collection and processing. Needless to say, for teaching purposes the processed data provided can also be used as input to any appropriate set of macromolecular crystallography programs in order to attempt the structure determination in the conventional way step by step.

In the following, some ideas are presented for carrying out the structure determinations and possibly modifying them for teaching purposes.

7.1. Insulin

S-SAD has become more and more popular in the past decade. Nevertheless, the very small signal that can be obtained from the anomalous scattering of light atoms still places such an experiment at the forefront of what is currently possible with good crystals and good data collection equipment. Cubic insulin is among the successful S-SAD cases, certainly one of the easier examples. Provided that the data are properly processed, the sulfur substructure can be determined straightforwardly. Because of the relatively high resolution of the data (1.6 Å), an interpretable map should be achieved quickly. In order to increase the requirements for good data processing, the redundancy of the data set can be decreased by taking only 90, 120 or 150° instead of the full 180° of data. In doing so, structure determination experiments of increasing difficulty can be created.

7.2. Thaumatin

Quick-soaking of protein crystals using highly concentrated halide solutions is a fast and easy way to produce heavy-atom

derivatives for phase determination (Dauter & Dauter, 1999). The thaumatin crystal used for data collection was soaked for a few seconds only, in order to limit the bromide penetration into the crystal. As a consequence, structure solution is successful only when two, three or four wavelengths of the four-wavelength MAD data sets are used. Again, various combinations of wavelengths may be tried to make up experiments of increasing difficulty. For more experienced students, it may even be possible to solve the structure based on a single data set (peak or inflection point).

7.3. Monoclinic lysozyme

When crystallizing monoclinic HEWL, one has to be aware of the fact that sometimes triclinic HEWL crystals also appear in the same experiment. These cannot be distinguished from the monoclinic crystals visually, but only when they are

mounted for data collection and indexed. The major difficulty in this project stems from the fact that the β angle of the crystal is close to 90° and that the b and c axes are almost equal in length. This makes the indexing of the diffraction images difficult. If this happens to be the case, different wedges of the data set may be tried for indexing. Structure solution by molecular replacement is straightforward, although high R values are obtained initially. The correctness of the solution can, however, easily be assessed by refinement. During the refinement process many triangular features will show up in the electron density maps, indicating surface-bound nitrate ions.

7.4. Tetragonal lysozyme (NaCl form)

The NaCl form of tetragonal HEWL is probably the most widely used crystal form for teaching in macromolecular

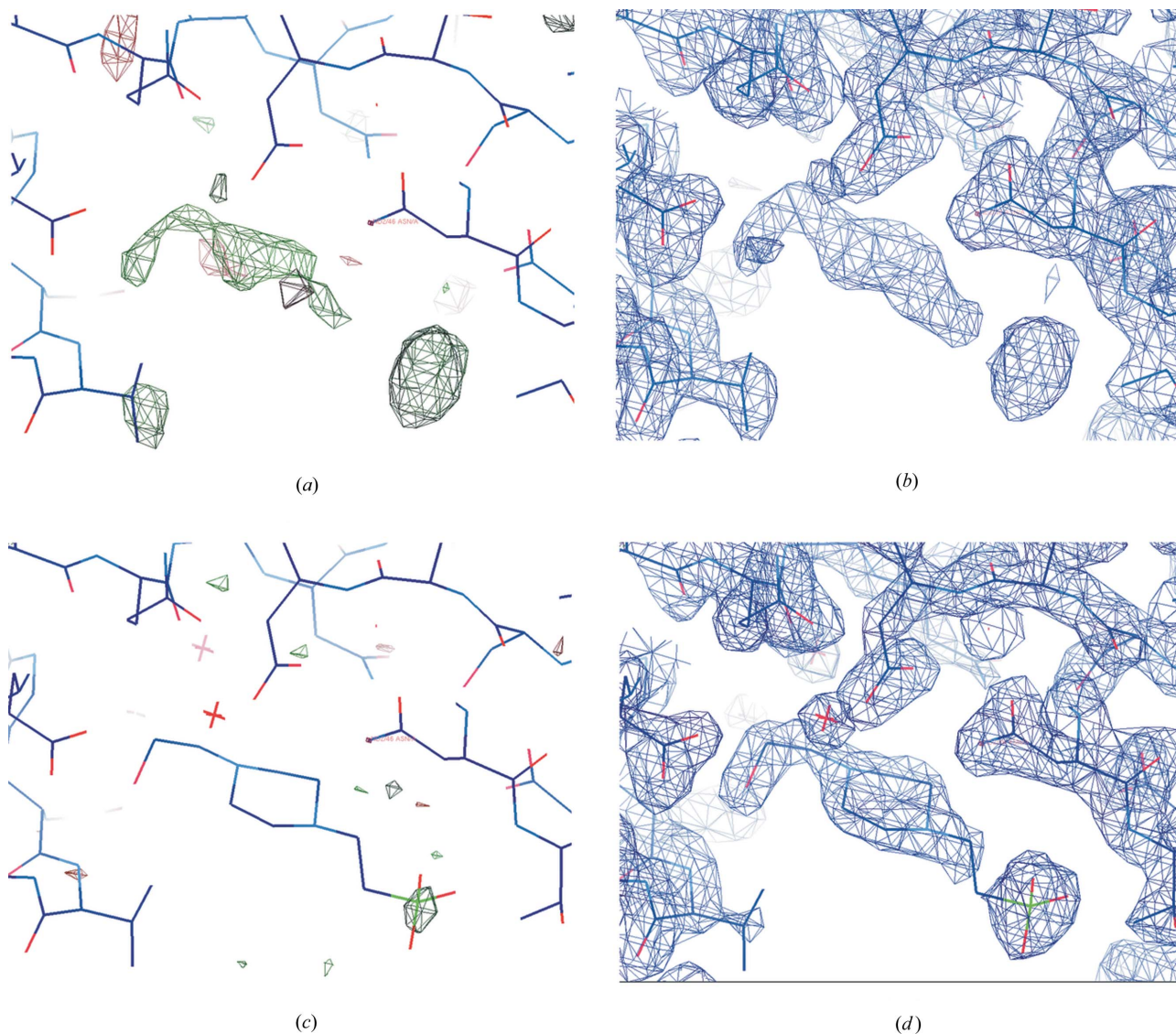


Figure 8 HEPES binding to HEWL. (a) ($F_{\text{obs}} - F_{\text{calc}}$, α_{calc}) electron density contoured at $+2.0\sigma$ (green) and -2.0σ (red); (b) ($2F_{\text{obs}} - F_{\text{calc}}$, α_{calc}) electron density contoured at 1.0σ . In both cases the HEPES model was not included in the phase calculation. (c) Anomalous difference Fourier electron density map contoured at $+3.0\sigma$, exhibiting 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. (d) Final ($2F_{\text{obs}} - F_{\text{calc}}$, α_{calc}) electron density contoured at 1.5σ for the final model including the HEPES ligand.

crystallography. However, one difficulty here is that the crystals tend to grow very large, which is nice from an aesthetic point of view, but which makes shock-cooling using paraffin oil as a cryoprotectant tricky. The crystals are often damaged and exhibit large diffraction spots and high mosaicity. It is therefore advisable to choose smaller crystals for data collection. Apart from that, the experiment is straightforward and poses no problems in data collection and processing. One aspect to be aware of is that the goal of the experiment is to obtain anomalous signals from light atoms. Therefore, good data quality is essential. This is achieved by collecting a full revolution (360°) around the rotation axis, which results in a highly redundant data set. Should a three- or four-circle diffractometer or kappa-geometry be available, the experiment may be modified by applying different data collection strategies. Rotating the crystal about more than one axis will result in higher data quality and therefore less redundancy will be required to achieve the same signal-to-noise ratio. If the data quality turns out to be exceptionally high, experimental phase determination may be attempted, similar to experiment 1.

7.5. Tetragonal lysozyme (HEPES/MPD form)

This crystal form of HEWL is a little harder to grow than the NaCl form and it grows less reproducibly. Furthermore, the crystals are often grown together, as shown in Fig. 3(e). Other than that, they pose no significant difficulty. The major challenge of this experiment is to identify the ligand, which is weakly bound and probably not fully occupied. This is of course a situation that researchers are often faced with in the pharmaceutical industry, where the binding mode of lead compounds to their target protein may have to be identified although the binding may still be weak. After carefully refining the protein structure the model phases are good enough to reveal residual electron density in the ($F_{\text{obs}} - F_{\text{calc}}$, α_{calc}) map. Additional support for the binding comes from inspecting the ($2F_{\text{obs}} - F_{\text{calc}}$, α_{calc}) and the (ΔF , $\alpha_{\text{calc}} - 90^\circ$) maps as well. The occupancy of the ligand may be estimated by comparing the peak height of the protein S atoms in the anomalous difference electron density map with the peak height of the S atom of HEPES.

8. Other courses and tutorials

The tutorial presented here is not the only one existing. A number of well known courses and workshops for macromolecular crystallography have been designed around the world. Many of them are run on a regular basis. Just to name a few, there is the M2M course in Hamburg, the RapiData course in Brookhaven and the Cold Spring Harbor Crystallography School. However, all these have in common that only a limited number of students is accepted and that course material beyond lecture notes is not made publicly available. Some tutorials also exist on the Internet, such as, for instance, Bernhard Rupp's *Crystallography 101* (<http://www.ruppweb.org/Xray/101index.html>) or Michael Sawaya's tutorial at <http://www.doe-mpi.ucla.edu/people/sawaya/>. A very good and

useful compilation of such tutorials is given on the American Crystallographic Association web site (<http://www.americalcrystallographic.org/CrystalEducationSites.html>). However, these tutorials are geared towards giving an overview of X-ray crystallography as a technique, rather than presenting specific examples to students for self-learning. Furthermore, they do not provide the diffraction data, let alone the raw diffraction images. Of course, diffraction data are available from the PDB (<http://www.rcsb.org/pdb>) and also, for instance, from the Autostruct web site (<http://www.ccp4.ac.uk/autostruct/testdata/index.html>) hosted by CCP4 or from the Joint Center of Structural Genomics (JCSG; <http://www.jcsg.org/datasets-info.shtml>). With the possible exception of the JCSG data, the majority of the data available from these sites are processed data. The few cases where raw diffraction images are available are left rather uncommented. Tutorials for more specialized applications in macromolecular crystallography or for a certain suite of computer programs are also available on the Web. To name them all would be beyond the scope of this contribution.

With the here-presented tutorial describing (i) an S-SAD structure determination of cubic insulin, (ii) a Br-MAD structure determination of thaumatin, (iii) a structure determination by molecular replacement using monoclinic lysozyme, (iv) the identification of bound solvent ions in lysozyme using a longer-wavelength data set and (v) the identification of a weakly bound ligand at the active site of lysozyme, we attempted to cover a broad range of macromolecular crystallography experiments. We certainly do not want to compete with any of the courses or tutorials mentioned above, but to the best of our knowledge this tutorial is unique in that it provides a complete description of a crystallographic experiment covering all aspects from crystallization, crystal mounting, data collection and processing to structure determination and interpretation. It provides not only the crystallization recipes but also the sources of all relevant materials, as well as detailed instructions to perform the experiments. In addition, example diffraction data are provided for those who do not have easy access to a diffraction facility, together with information on the various steps from data processing to structure determination. All of the computer programs used and described in the tutorial are freely available for academic users.

9. Availability of the data and the tutorial

The entire tutorial including the raw diffraction images for all five experiments, the integrated and scaled data, the solved and refined structures, and detailed descriptions of the data processing and structure solution are available as supplementary information to this paper.¹ All data can also be downloaded from the sites http://www.embl-hamburg.de/Xray_Tutorial and <http://www.mx.bessy.de/bessy-ws/bessy.html>

¹ Supplementary data for this paper are available from the IUCr electronic archives (Reference: KK5029). Services for accessing these data are described at the back of the journal.

and are available from the authors upon request. The tutorial has been successfully used during the workshop for which it was designed and created (see <http://www.embl-hamburg.de/workshops/2007/diffraction>). Furthermore, it has been used and favourably received by about 20 individual researchers and students to date, who have learned about it by word of mouth or from presentations at the 16th Annual Meeting of the German Society of Crystallography and the XXI Congress and General Assembly of the International Union of Crystallography (Faust *et al.*, 2008).

We acknowledge the support of the DGK (German Society of Crystallography) for their financial contribution towards a workshop during which the data were collected and the beamline staff at the BESSY synchrotron in Berlin-Adlershof (Germany) for their help in organizing the workshop.

References

- Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N. & Bourne, P. E. (2000). *Nucleic Acids Res.* **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 (London)*, **206**, 757–761.
- Brinkmann, C., Weiss, M. S. & Weckert, E. (2006). *Acta Cryst.* **D62**, 349–355.
- Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-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.
- Dauter, Z. & Dauter, M. (1999). *J. Mol. Biol.* **289**, 93–101.
- Dauter, Z., Dauter, M., de La Fortelle, E., Bricogne, G. & Sheldrick, G. M. (1999). *J. Mol. Biol.* **289**, 83–92.
- Diederichs, K. & Karplus, P. A. (1997). *Nat. Struct. Biol.* **4**, 269–275.
- Djinović Carugo, K., Helliwell, J. R., Stuhmann, H. & Weiss, M. S. (2005). *J. Synchrotron Rad.* **12**, 410–419.
- Emsley, P. & Cowtan, K. (2004). *Acta Cryst.* **D60**, 2126–2132.
- Evans, G. & Pettifer, R. F. (2001). *J. Appl. Cryst.* **34**, 82–86.
- Faust, A., Panjikar, S., Mueller, U., Parthasarathy, V., Schmidt, A., Lamzin, V. S. & Weiss, M. S. (2008). *Acta Cryst.* **A64**, C80.
- Hao, Q. (2004). *J. Appl. Cryst.* **37**, 498–499.
- Hendrickson, W. A. (1999). *J. Synchrotron Rad.* **6**, 845–851.
- Hendrickson, W. A. & Ogata, C. M. (1997). *Methods Enzymol.* **276**, 494–522.
- Hendrickson, W. A. & Teeter, M. M. (1981). *Nature (London)*, **290**, 107–113.
- Jaskólski, M. (2001). *J. Appl. Cryst.* **34**, 371–374.
- Kabsch, W. (1993). *J. Appl. Cryst.* **26**, 795–800.
- Ko, T.-P., Day, J., Greenwood, A. & McPherson, A. (1994). *Acta Cryst.* **D50**, 813–825.
- 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.
- Mueller-Dieckmann, C., Panjikar, S., Tucker, P. A. & Weiss, M. S. (2005). *Acta Cryst.* **D61**, 1263–1272.
- 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). *Nat. Struct. Biol.* **6**, 458–463.
- Saraswathi, N. T., Sankaranarayanan, R. & Vijayan, M. (2002). *Acta Cryst.* **D58**, 1162–1167.
- Schneider, T. R. & Sheldrick, G. M. (2002). *Acta Cryst.* **D58**, 1772–1779.
- Sheldrick, G. M. (2002). *Z. Kristallogr.* **217**, 644–650.
- Sheldrick, G. M. (2008). *Acta Cryst.* **A64**, 112–122.
- Sheldrick, G. M., Hauptman, H. A., Weeks, C. M., Miller, R. & Uson, I. (2001). *International Tables for Crystallography*, Vol. F, edited by M. G. Rossmann & E. Arnold, pp. 333–351. Dordrecht: Kluwer Academic Publishers.
- Vagin, A. & Teplyakov, A. (1997). *J. Appl. Cryst.* **30**, 1022–1025.
- Vaney, M. C., Maignan, S., Riès-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 ed. Hoboken: John Wiley and Sons Inc.
- Weiss, M. S. (2001). *J. Appl. Cryst.* **34**, 130–135.
- Weiss, M. S., Palm, G. J. & Hilgenfeld, R. (2000). *Acta Cryst.* **D56**, 952–958.
- Weiss, M. S., Sicker, T. & Hilgenfeld, R. (2001). *Structure*, **9**, 771–777.