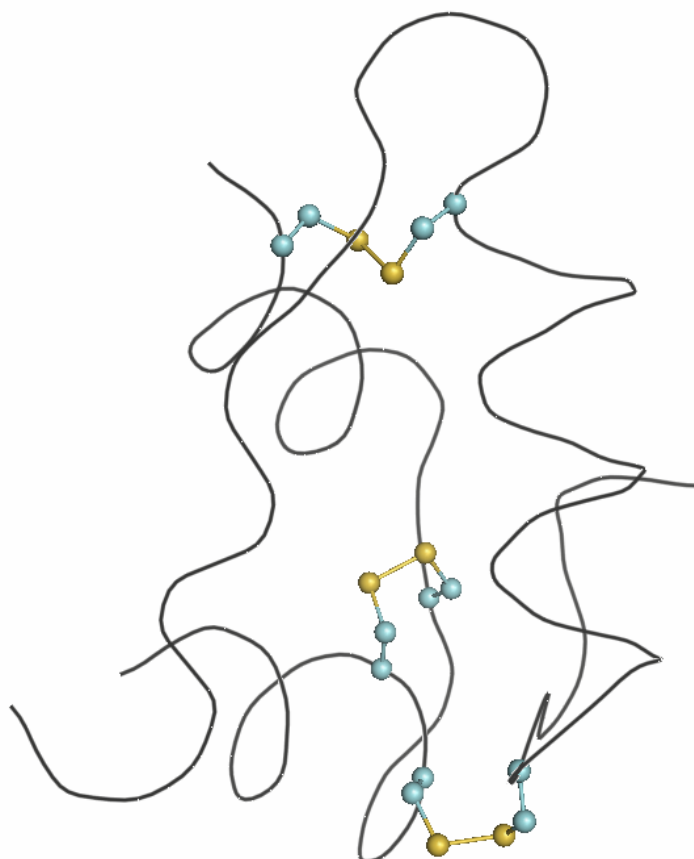


A tutorial for learning and teaching macromolecular crystallography

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



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

Experiment 1: S-SAD on bovine Insulin

Insulin regulates the cellular uptake, utilisation, 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 ($M_w = 5,800$ Da). The amino acid sequence is given in Figure 1. Three disulfide bonds hold the two chains together, one intra-chain SS-bridge between Cys6 and Cys11 in chain A and two interchain 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.

Experimental phase determination using single-wavelength anomalous diffraction (SAD) from the sulphur 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, 6 S-atoms) by Hendrickson and Teeter (1981), it took 18 years until the method was rediscovered by Dauter and colleagues (Dauter *et al.*, 1999), who were able to demonstrate that the structure of hen egg-white lysozyme (HEWL) could be successfully determined from the anomalous scattering of the protein S-atoms and surface-bound chloride ions alone. Since then, approximately 100 structures have been obtained using this so-called sulphur-SAD or S-SAD approach. Since 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 the diffraction data collection at somewhat longer wavelengths may be beneficial (Djinovic-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 $\lambda = 1.77$ Å.

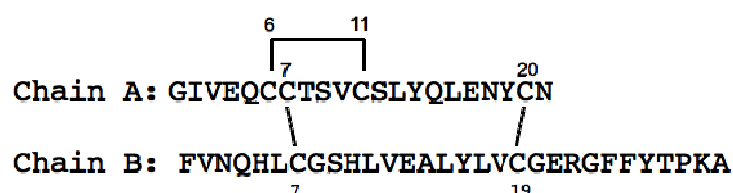


Figure 1: Amino acid sequence of bovine insulin including the 3 disulfide-bonds

1 Crystallisation

Chemicals:	bovine insulin	(M= 5733.49 g/mol, Sigma cat.no. I5500)
	Na ₂ HPO ₄ *12H ₂ O	(M= 358.14 g/mol, Fluka cat.no. 71649)
	Na ₃ PO ₄ *12H ₂ O	(M= 380.12 g/mol, Fluka cat.no. 71908)
	Na ₄ EDTA*4H ₂ O	(M= 452.23 g/mol, Fluka cat.no. 03699)
	Na ₂ EDTA*2H ₂ O	(M= 372.24 g/mol Fluka cat.no. 03679)
	Glycerol	(M= 92.09 g/mol, Sigma-Aldrich cat.no. G9012)
	Milli-Q water	
	EasyXtal Tool screw-cap crystallization plates	(Nextal, now Qiagen)

Bovine insulin crystals were prepared by the hanging drop-method in EasyXtal Tool crystallization plates. 4 µl of protein solution (20 mg/ml of protein dissolved in 20 mM Na₂HPO₄ and 10 mM Na₃EDTA pH 10.0–10.6) and 4 µl of reservoir solution (225–350 mM Na₂HPO₄/Na₃PO₄ pH 10.0–10.6, 10 mM Na₃EDTA) were mixed and equilibrated over reservoir solution. Crystals belonging to the cubic space group I2₁3 (space group number 199) with the unit-cell parameter $a=78.0 \text{ \AA}$ grew within a few days up to a final size of 100–300 µm³. They were cryo-protected in a solution containing 250 mM Na₂HPO₄/Na₃PO₄ pH 10.2, 10 mM Na₃EDTA, and 30% (v/v) glycerol, and usually diffracted X-rays to better than 1.4 Å.

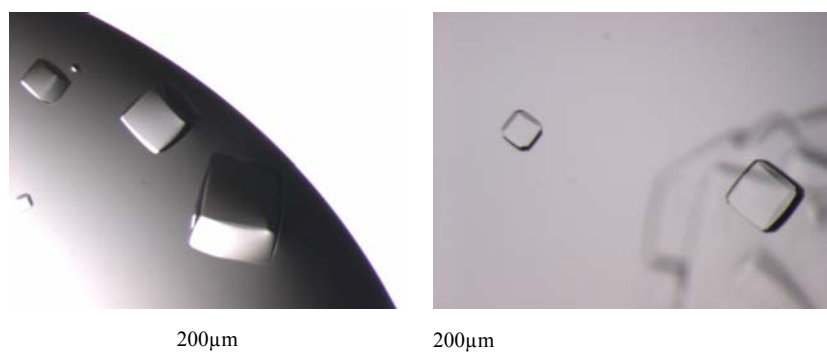


Figure 2: Cubic insulin crystals

2 Data collection

Diffraction data were collected at a wavelength of 1.77 Å at the tunable beam line BL 14.1 at the BESSY synchrotron in Berlin Adlershof. BL 14.1 is equipped with a MARMosaic-CCD detector (225mm) from the company MARRESEARCH (Norderstedt, Germany) and a MARdtb goniostat (MARRESEARCH, Norderstedt, Germany).

The relevant data collection parameters are given below:

wavelength:	1.77 Å
detector distance:	50 mm
2 θ -angle:	0°
oscillation range/image:	1.0°
no. of images:	180
path to images:	experiment1/data
image name:	ins_ssad_1_###.img
exposure time/image:	2.6 sec

Based on the chemical composition of the insulin crystals and the tabulated anomalous scattering lengths, the expected Bijvoet ratio as a function of the data collection wavelength can be calculated (Figure 3). At the data collection wavelength chosen, it is about 1.7%

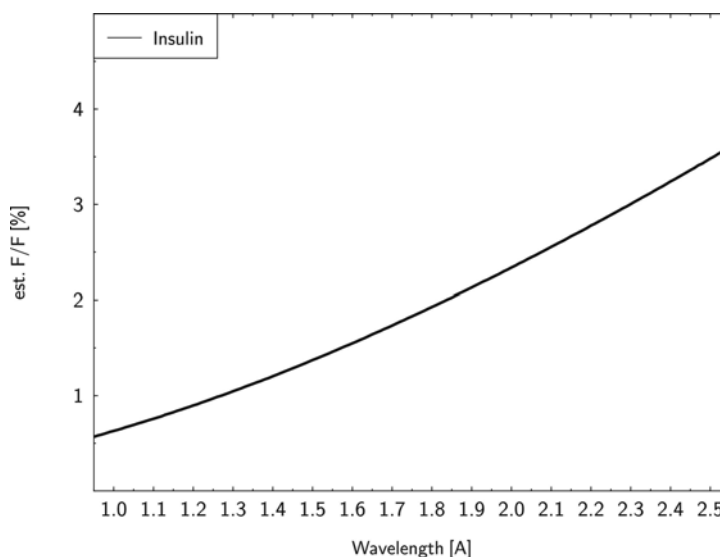


Figure 3: Estimated $\Delta F/F$ for Insulin as a function of data collection wavelength. The chemical composition used was $C_{255} H_{376} N_{65} O_{75} S_6$.

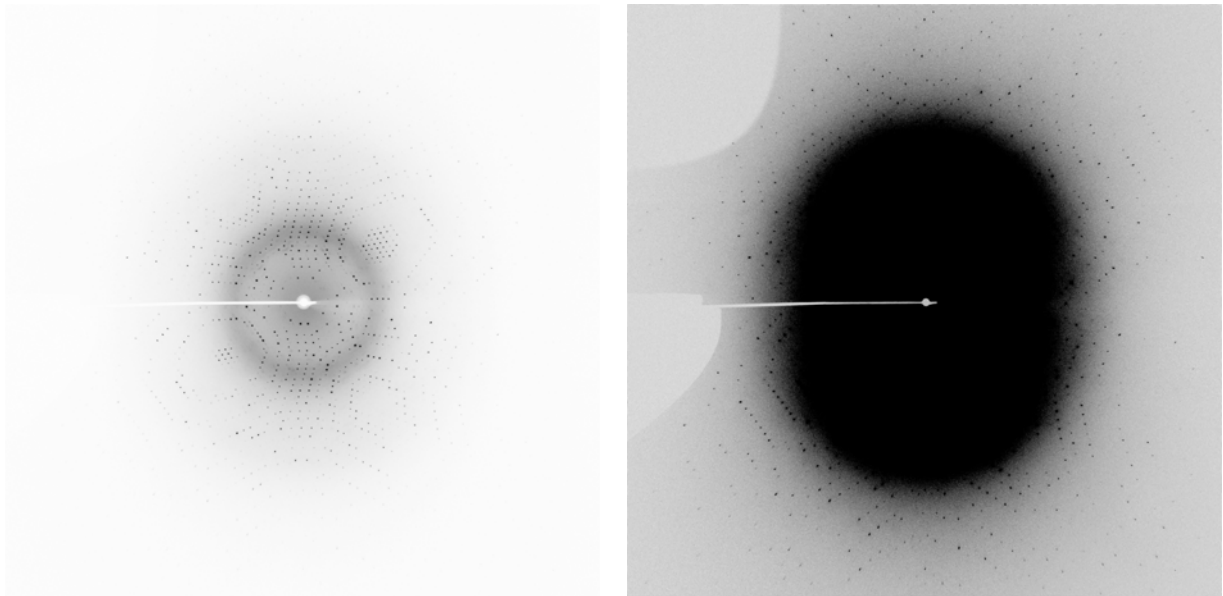


Figure 4: Diffraction image of cubic insulin displayed at different contrast levels. The shadow in the upper left hand corner on the image on the right originates from the cryo-nozzle. The other shadow on the left side is caused by the beam stop and its holder.

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 is recognized by the program. In XDS.INP the image name given must not include the zipping-format extension (*.img instead of *.img.bz2). Further, XDS has a very limited string length (80) to describe the path to the images. Therefore it may be necessary to create a soft link to the directory containing the images by using the command **ln -s /path/to/images/ ./images**. The path to the images in XDS.INP will then be ./images/.

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

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

JOBS= XYCORR INIT COLSPOT IDXREF space group number=0

XYCORR	computes a table of spatial correction values for each pixel
INIT	determines an initial background for each detector pixel and finds the trusted region of the detector surface.
COLSPOT	collects strong diffraction spots from a specified subset of the data images.
IDXREF	interprets observed spots by a crystal lattice and refines all diffraction parameters.

The IDXREF.LP output file contains the results of the indexing. For insulin the correct space group is I2₁3 (space group number 199) with cell parameter a = 78.0 Å.

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

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

DEFPIX	defines the trusted region of the detector, recognizes and removes shaded areas, and eliminates regions outside the resolution range defined by the user.
XPLAN	helps planning data collection. Typically, one or a few data images are collected initially and processed by XDS. XPLAN reports the completeness of data that could be expected for various starting angles and total crystal rotation.

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

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

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

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

- **scaling** **run XSCALE**

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

Table 1. Data processing statistics (from XSCALE.LP). The numbers in parentheses refer to the outermost resolution limit.

Resolution limits [Å]	50.0-1.60 (1.70-1.60)
Space group	I2 ₁ 3
Unit cell parameters a, b, c [Å]	78.0
Mosaicity [°]	0.14
Total number of reflections	208,859
Unique reflections	20,226
Redundancy	10.3 (8.3)
Completeness [%]	99.9 (100.0)
I/σ(I)	48.1 (13.6)
R_{r.i.m.} / R_{meas} [%]	3.7 (17.0)
Wilson B-factor	22.1

- **converting *.ahkl to *.mtz**

run XDSCONV with XDSCONV.INP

XDSCONV.INP:

```
OUTPUT_FILE=ssad_insulin.mtz CCP4  
INPUT_FILE=ssad_insulin.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 ssad_insulin_ccp4.mtz << EOF  
LABIN FILE 1 ALL  
END  
EOF
```

Alternatively, the XDS_ASCII.HKL file can be converted with COMBAT and the resulting mtz-file can be used as input file for scaling in the CCP4-program SCALA.



Figure 5: Anomalous difference Fourier electron density map with the six heavy atoms sites from SHELXD. The map is contoured at 8σ .

The correct hand for the substructure was determined using the programs ABS (Hao, 2004) and SHELXE (Sheldrick, 2002). Initial phases were calculated after density modification using the program SHELXE and extended to 1.60 Å resolution. 90.2% of the model was built using the program ARP/wARP 7.0 (Perrakis *et al.*, 1999; Morris *et al.*, 2002). More details can be found in the attached AUTO-RICKSHAW output (directory experiment1/autorickshaw). The complete Auto-Rickshaw run in the advanced version took around 20 minutes. The model was then further modified using COOT (Emsley, 2004) and refined using REFMAC5 (Murshudov *et al.*, 1997). Figures 6 and 7 show snapshots of the final model superimposed with the anomalous difference map and the experimental electron density map after density modification in DM.

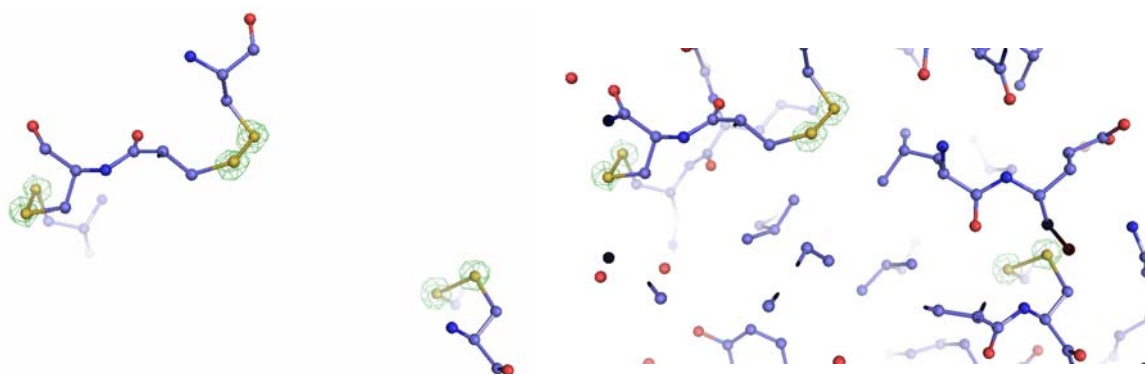


Figure 6: Final model superimposed with the anomalous difference Fourier electron density map. Left panel: the disulfide bridges in the final model; Right panel: a larger part of the final model. The map is contoured at 8σ .

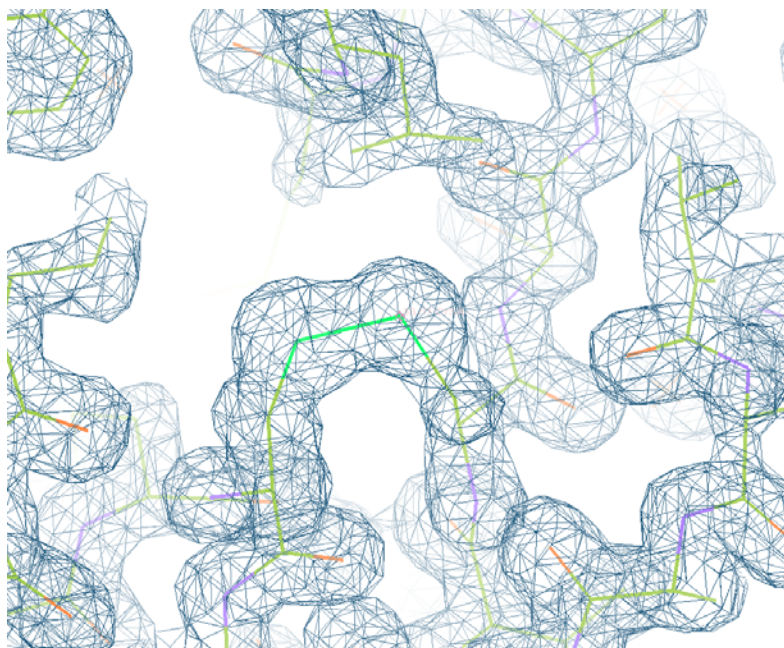


Figure 7: Experimental electron density map after solvent flattening using the program DM superimposed onto the final refined model. The map is contoured at 1.5σ .

5 References

- Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* **D50**, 760-763.
- Cowtan, K. (1994). *Joint CCP4 and ESF-EACBM Newsletter on protein crystallography* **31**, 34-38.
- Dauter, Z., Dauter, M., de La Fortelle, E., Bricogne, G. & Sheldrick, G. M. (1999). *J. Mol. Biol.* **289**, 83-92.
- 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.
- Hao, Q. (2004). *J. Appl. Cryst.* **37**, 498-499.
- Hendrickson, W. A. & Teeter, M. M. (1981). *Nature* **290**, 107-113.
- 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., Tucker, P. A. & Weiss, M. S. (2005). *Acta Cryst.* **D61**, 1263-1272.
- Murshudov, G. N., Vagin, A. A. and 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.
- Schneider, T. R. & Sheldrick, G. M. (2002). *Acta Cryst.* **D58**, 1772-1779.
- Sheldrick, G. M., Hauptman, H. A., Weeks, C. M., Miller, R. & Uson, I. (2001). *International Tables for Macromolecular Crystallography, Vol. F, edited by M. G. Rossmann & E. Arnold*, **ch. 16**, pp. 333-345. Dordrecht: Kluwer Academic Publishers.
- Sheldrick, G. M. (2002). *Z. Kristallogr.* **217**, 644-650.
- Sheldrick, G. M. (2008). *Acta Cryst.* **A64**, 112-122.
- 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., Sicker, T. & Hilgenfeld, R. (2001). *Structure* **9**, 771-777.