ESRF	**Experiment title: Structural Studies of Thi1, a protein involved in DNA repair and thiamine biosynthesis**	**Experiment number**: LS−1909
Beamline: ID29	**Date of experiment**: from: 13 july 2001 to: 14 july 2001	**Date of report**: 22 february 2002
Shifts: 3	**Local contact(s)**: Dr. Bill Shepard	*Received at ESRF:*

Names and affiliations of applicants (* indicates experimentalists):

Prof. Dr. Glaucius Oliva

Paulo Henrique Conaggin Godoi *

Institute of Physics in São Carlos and
Institute of Chemistry in São Carlos;
University of São Paulo;
Brazil.

Report:

Crystals from two projects have been used for data collection: Thi1 from *A. thaliana* and N−acetylglucosamine−6−Phosphate Deacetilase from *E. coli* (EC 3.5.1.25).

The *Arabidopsis thaliana thi1* gene was cloned by functional complementation of *E. coli* mutant strain BW535, defective in DNA base excision repair pathways. This gene partially complements the methyl−methane sulfonate sensitive phenotype of BW535 and partially complements the UV sensitive phenotype of *E. coli* AB1886. Its cDNA correspond to an open reading frame of 1047 bp whose predicted amino acid sequence (36kDa) has homology to genes related to the thiazole biosynthesis but, actually, to no other enzyme clearly involved in DNA repair. In spite of it, the predicted amino acid sequence revealed protein motifs which are consistent to its expected activity: a dinucleotide binding site and a bacterial DNA polimerase active site.

The protein was expressed and purified from *E. coli* as a 6xHis−tagged protein. This procedure allowed the purification of high quantities of protein samples in a high purity degree. On the other hand, the 6xHis−tag sequence have been seen to be inhibitory to cristallization. Prior to the crystallization experiments, the Thi1 protein was digested with papain and purified. Crystals were grown in 100mM MES pH6.0, 40% MPD and 1.5% (w/v) 1,2,3−heptanetriol with a maximum size in the order of 100 x 100 x 10 μm. For the experiments in ESRF, 18 native crystals were visually selected, mounted in nylon cryolooops and flash−frozen in a 100 K nitrogen stream. Other 4 crystals were also prepared by soaking in $HgCl_2$, K_2PtCl_4 and $KAuCl_4$ and transported in a dry−shipper dewar to ESRF.

The established priority to the ESRF data collection was to obtain the best native data set following to the data collection of the potential derivatives. Once in the beam line, the native crystals were inspected. The one diffracting to the higher resolution, with the lowest mosaicity and best signal−to−noise statistics was collected with high redudancy. Among the derivative crystals, the platinum derivative was the only useful crystal collected. A wide−fluorescence scan confirmed the presence of a clear absorbance edge which is characteristic of Pt and other four unexpected edges. One of these, as intense as the Pt signal, was recognized as a characteristic Zn border. Since zinc ions were not added in any of the purification steps, neither in the crystallization conditions, the only avaliable source for zinc ions shall be provenient of the LB growth medium used for the protein expression. Knowing that this zinc ion is certainly an intrinsic scatterer, the zinc signal was taken as the first priority to the MAD data collection.

The crystals belong to the space group F222 with unit cell parameters of a=102.4, b=133.3 and c=142.6Å. A five−wavelenght data set were collected by cryocrystallography using the best diffracting crystal crystal in the Zn−peak wavelenght (1.28269 Å), Zn−edge (1.28321 Å), Pt−peak (1.07232 Å), Pt−edge (1.07272 Å) and in the hard−remote (0.915013 Å). The crystal decay was considered small during the data−collection. The images were analysed and the reflexions were integrated and indexed by DENZO and SCALEPACK. The data collection statistics is as follows:

Data sets −−−−−−>	native	Zn−peak	Zn−edge	Pt−peak	Pt−edge
Wavelenght (Å)	0.915013	1.28269	1.28321	1.07232	1.07272
Resolution limits (Å)	40−1.60	40−2.35	40−2.35	40−2.0	40−2.0
Completeness (%)	99.4 (94.3)	97.9 (96.3)	97.5 (95.1)	97.5 (95.6)	97.6 (95.3)
Redundancy = 4 (% data)	81.1	89.2	59.6	62.2	61.9
$\langle I \rangle / \langle \sigma(I) \rangle$	16.5 (1.3)	24.4 (7.8)	21.2 (6.8)	16.1 (3.0)	17.27 (3.1)
R_{sym} (%)	5.7 (22.4/49.8)*	3.6 (9.6)	4.4 (12.0)	5.5 (28.8)	5.0 (26.6)

The values in parentheses are those found in the highest resolution shell. *here the values refers to data collected to 1.6 and 1.8 Å resolution respectively.

On the light of these results, we calculated the Mattews coefficient as 3.8, considering 2 molecules in the assymetric unit and a solvent content of 34.4%. The fast self−rotation function calculated by GLRF confirmed the presence of a non−crystallographic symmetry axis parallel to the ac plane.

The positions of the anomalous scatterers and initial phasing were performed by the SOLVE/RESOLVE program, treating the data as a SAD case (in the maximum δf" contribution, *i.e.* the Zn−peak or Pt−peak data sets), as a single MAD data set (Zn or Pt borders) and as a multiple MAD data set. Unfortunately, the results indicate that only the Zn−peak data set contains a small but significant signal−to−noise statistics for phasing. The Zn−peak data set is now being used in phasing, density modification and model building.

Two Zn atoms were found in the assymetric unit (in fractional coordinates): for the Zn1 atom x=0.5611, y=0.1526, z=0.1937, occupancy=0.36 and B=24.1 $Å^2$; for the Zn2 atom x=0.8269, y=0.1525, z=0.1144, occupancy=0.41 and B=33.8 $Å^2$. The mean figure−of−merit for this solution is only 32% using all data between 20 and 3.0 Å. We will now include NCS averaging during the next phasing refinement, following to the solvent flattening procedure.

In concern to the data collection of N−acetylglucosamine−6−Phosphate Deacetilase from *E. coli*, two potential derivative crystals were soaked in $KAuCl_4$ and measured at λ=0.915013 Å. At this point, the theoretical values of f' and f" are −8.861 and 8.167 e^-, respectively. The data set was processed by MOSFLM and SCALA and it was seen that the anomalous signal present is of poor phasing power.