FIP	**Rapport d'expérience** **Utilisation du temps de faisceau** **Run 2002/1**	BM30A

Identification de l'expérience date début-fin **8-12 novembre 2002**

n° proposition **30-01-556**

nom du laboratoire ou nom du responsable du projet **LEBS-CNRS / Gif sur Yvette**

nom des utilisateurs présents

Marc Graille, Ronald Melki, **Nicolas Fay**, **Luc Bousset, Sebastiano Pasqualato, Louis Renault,** LEBS-CNRS, 1 Av. de la Terrase, Gif sur Yvette, France

Nicolas Leulliot, LURE, Bat. 209D, centre universitaire Paris Sud BP34, 91898 Orsay Cedex, France

Agnidipta Ghosh, Max-Planck Institut of Dortmund, Postfach 50 02 47, 44 202 Dortmund, Germnay

nom du local contact **Lilian Jacquamet**

Temps de faisceau mode / intensité **16 Bunch**

temps alloué **12 shift**

temps utilisé **11.5 shift**

Statistique d'utilisation

MAD	oui [X]	seuil:	non [**X**]	$\lambda = 0.980$
cryo	oui [**X**]		non [**X**]	

Rapport d'expérience et commentaires

Marc Graille, Nicolas Leulliot, Herman Van Tilbeurgh (6 shifts on 8-10/11/02 and 2 shifts on 14-15/09/02): Yeast structural genomics project

We have used 2 shifts on FIP on 14-15th September 2002 and 6 shifts on 8-10th November 2002 for collecting MAD data sets on different yeast structural genomics projects (Sophie Quevillon-Cheruel et al. A structural-genomics initiative on yeast proteins. *Accepted in J of Synchr radiation*) and few native data sets on other projects. More details on every orf of yeast structural genomics projects can be found on http://genomics.eu.org/targets.html. All the statistics from data collection, reduction and scaling for the use of FIP beam time are grouped together and summarised in table 1.

1) YML079w (target 136)

This orf of unknown function crystallises as thin plates. Crystals suffer from severe anisotropy but useful data could be collected on native crystals during this run. Unfortunately, anisotropy gets beyond acceptable levels for the Se-Met substituted crystals. Efforts are being oriented towards improvements of the quality of the Se-Met crystals.
A few data sets on Se-Met crystals were obtained, but these data did not allow to find the SE positions.

2) YGR205w (target 43)

This orf codes for a protein of unknown function. A P-loop motif was recognized in the sequence. This feature is a strong signature for nucleotide fixing. Native crystals were obtained for the protein and the structure was solved by MAD data collection at three different wavelengths to a resolution of 2.8A. The structure was refined against native data at 2.1A resolution. The structure revealed an unexpected strong resemblance to the structure of pantothenate kinase from *E. coli*, although sequence analogy between both proteins is beyond detectable levels. No pantothenate activity could be detected in the enzyme in our hands and therefore the function of this protein remains unknown. We could biochemically prove the binding of ATP to the enzyme. Considerable effort was put into the cocrystallisation of Orf43 in the presence of ATP or ATP analogues. A number of data sets was collected on these crystals. Unfortunately ATP could never be observed in the electron density. However, a sulphate ion from the crystallisation mother liquor is strongly bound to the P-loop and this suggests a possible binding mode for at least the phosphate group of nucleotides.

3) YER010c (target 154)

This protein is of unknown structure and function. Crystals of the Se-Met substituted protein were obtained under the same crystallization conditions as the native protein. Analysis of the fluorescence scan spectrum revealed that Se was partially oxidised in the crystals. Data were therefore collected at 4 different wavelengths. The protein contains only two methionines for 234 residues and the corresponding anomalous signal was too weak to determine protein phases. Site directed mutagenesis was carried out to incorporate more methionines into the protein. A triple methionine mutant was produced and the native as well as the Se-met substituted versions of this mutant were crystallised. Native and complete MAD data were collected on these crystals. However, the presence of cacodylate in the crystallization buffer masked the fluorescence spectrum and the ensuing data contained insufficient signal to position the Se atoms in the crystal. Crystals are now prepared that do not contain cacodylate buffer.

4) YDR435c (target 182)

This orf codes for a N-terminal methyl transferase involved in the regulation of a phosphatase complex. A four wavelength data set was collected on the Se-met substituted crystals. The protein contains 14 methionines and the structure could be solved in an almost automatic mode.
Complete data sets on the native protein with and without a potential cosubstrate have also been collected to a resolution of 2.0A (100 % complete Rsym =6%). Difference electron density maps allowed to determine the structure of the bound SAM.
The structure is clearly organised as two structural domains: one has the classical methyl transferase β-sheet fold, and a second helical domain is more specific to the enzyme.

5) YOR357c (orf190)

This orf codes for a protein consisting essentially of a PX domain. PX domains have been recently identified in a number of different proteins. At the biochemical level it was shown that these modules of about 130 amino acids are binding to phosphoinositides. PX domains are involved in various cellular functions such as vacuolar targeting and membrane protein trafficking. The yeast genome codes for 17 proteins containing PX domains. One of these, Grd19p consists of only the PX domain with 30 extra residues at the N terminal. It is homologous to the functionally characterized human sorting nexin protein SNX3. Grd19P is involved in the localization of the late-Golgi membrane proteins DPAP A and Kex2p. The protein was crystallised as well as the Se-methionine substituted version. We determined the crystal structure of Grd19P in the free form and in complex with diC4PtdIns(3)P at a resolution of 2.0 and 2.3 A respectively. This is the first case of an intact PX containing protein where both free and ligand bound conformations are available. The ligand occupies a well defined positively charged binding pocket at the interface between the β-sheet and a helical parts of the molecule. The structure of the free and bound protein are globally similar but show some significant differences in a region containing a polyproline peptide and the early characterised membrane attachment site. In the ligand bound form this region closes in upon the ligand. This information shows that no major conformational changes occur upon ligand binding. The first thirty residues in the structure, which are not homologous to other PX domain containing proteins, are not visible in the crystal structure and are partially proteolysed.

Table 1: Summary of the data collection at FIP.

Project.	**Date.**	**Reso.**	**Rsym**	**Completeness.**	**Space group.**	**Unit cell parameters.**	**Datasets collected.**	**Comments.**
136	08/11/02	2,5	9,1%	97%	P212121	a=56,4 b=69,3 c=108,67	1 SAD dataset.	Analysis in progess.
136	08/11/02	1,8	16.00 %	97%	P212121	a=56,4 b=69,3 c=108,67	1 native dataset.	Phases missing.
43 SeM peak	08/04/02	2,8	7,8%	99,6%	P43212	a=b=65,2 c=140,6	MAD	Structure phased and refined.
43 SeM inflexion	08/04/02	2,8	6,2%	98,9%	P43212	a=b=65,2 c=140,6	MAD	Structure phased and refined.
43 SeM remote	08/04/02	2,8	5,9%	98,8%	P43212	a=b=65,2 c=140,6	MAD	Structure phased and refined.
43-ATP	08/11/02	2,25	5,5%	98,9%	P43212	a=b=65,2 c=140,6	1 native dataset.	Refined structure. Ligand not present.
190	09/02/02	2	6,5%	99,6%	P6122		2 native datasets (2A).	Structure solved.
190 peak	09/02/02	2,3	5,7%	97,1%	P6122		3 MAD datasets.	Structure solved.
190 inflexion	09/02/02	2,3	3,2%	96,9%	P6122		3 MAD datasets.	Structure solved.
190 remote	09/02/02	2,3	6,5%	95%	P6122		3 MAD datasets.	Structure solved.
154M3 peak	08/11/02	2,6	8%	97,1%	P21212	a=128,1 b=254,8 c=48,5	MAD dataset.	Anomalous signal too weak.
154M3 inflexion.	08/11/02	2,6	8,6%	97,1%	P21212	a=128,1 b=254,8 c=48,5	MAD dataset.	Anomalous signal too weak.
154M3 remote.	08/11/02	2,6	9,4%	97,2%	P21212	a=128,1 b=254,8 c=48,5	MAD dataset.	Anomalous signal too weak.
182+SAM peak	13/09/02	2	6,8%	97,2%	P65	a=b=112,7 c=162,7	MAD dataset	Structure solved. Refinement in progress.
182+SAM inflection	13/09/02	2	8,1%	97,2%	P65	a=b=112,7 c=162,7	MAD dataset	Structure solved. Refinement in progress.
182+SAM remote	13/09/02	2	6,9%	97,2%	P65	a=b=112,7 c=162,7	MAD dataset	Structure solved. Refinement in progress.

Proteins outside the structural genomics project.

1) Complex between a human VpreB receptor and its stromal ligand Galectin 1.

Space group P212121; Resolution = 2.1 A; Completeness = 99 % and Rsym = 7,7%.

VpreB is a receptor protein which is expressed on the surface of developing B-cells at the preB stage. There has been a longstanding effort in the purification and crystallization of this atypical receptor. Crystals of the VpreB receptor were obtained in presence of its natural stromal ligand : galectin 1, a protein of about 120aa belonging to the mammalian lectin family. Analysis of the solvent content of the crystals showed that the whole complex could not be present in the asymmetric unit. The structure was solved by Molecular Replacement showing that indeed only the galectin1 crystallized. The structure is refined and is very similar to previous determined mammalian galectin structures. Studies are underway to characterize biochemically the interaction between VpreB and galectin in order to help us to get crystals of the complex.

2) Colicin D from *E. coli*.

Space group P41212 (a=62.8, c=148.7); Resolution =2.0 A; Completion = 99.8%; Rsym =4%.

Colicins are toxins that are used by bacteria to protect themselves against other microorganisms. They are secreted in the culture medium as a complex with the immunity protein, which protects the producing organism against toxic effects. Colicin D has a tRNA hydrolysing toxic activity. It was shown that ColD is proteolytically cleaved during entrance in the target cell. ColD has no sequence identity with colicins of known structure. We obtained crystals of the complex of the *E. Coli* ColD and its corresponding immunity protein. Native data have been collected at 2.6A and improvement of the crystals has now yielded data at 2.0A resolution. Biochemical analysis of the crystals revealed that they contain a proteolytic fragment of ColD that is generated during the crystallization process. Mass spectrometric analysis shows that probably the catalytic domain in complex with the immunity protein crystallized. Se-Met production of these fragments is on the way to solve the structure of the fragment. However Se-met versions of the protein fragments refuse to crystallize for the moment.

2) The transcriptional antiterminator LicT

Spacegroup $P3_221$; a=b=48A; c=166A; Resolution: 2.8A.

Se data:	completion (%)	Rsym (%)
Peak	88	3.4
Inflection	90	4.3
Remote	80	3.6

LicT is involved in the regulation of polyglucane degradation in *Bacillus subtilis*. LicT functions as a transcriptional antiterminator protein that consists of 2 structural and functional domains: one RNA binding domain at the N-terminus (55 residues) and one regulatory domain: PTS regulatory domain or PRD (220 residues). The protein is regulated through phosphorylation at well characterized histidine residues at the PRD domain. The phosphorylation state of these residues will determine the affinity for the antiterminator region in the transcribed mRNA region. The phosphorylated histidines can functionally be mimicked by mutation into aspartates. The doubly mutated H207D/H269D mutant is a

constitutively active form of the protein. We published its structure about one year ago. The structure consists of two very similar helical structural domains (PRD1 an PRD2) that are forming a tight dimer. This dimer buries the activating mutations, which are completely inaccessible to phosphorylating partners.

We determined the structure of the inactive wild type protein in order to characterize the conformational changes induced upon activation of the protein. Crystals were obtained of the wild type PRD and data collected in 2000 on FIP. Molecular replacement using the activated structure allowed us only to position the first module of PRD (PRD1). The structure of PRD2 (carrying the activating mutations) apparently diverged too much to be a useful model in molecular replacement. Obtaining crystals of the Se-Met substituted version of the wild type PRD was a very tedious task. We obtained one reasonably diffracting crystal which was used to collect a complete data set at three different wavelengths. Phases were useful to a resolution of 3.4A. Phases were extended using the program RESOLVE. The resulting maps were of poor quality but allowed to

1) confirm the molecular replacement solution for PRD1
2) position a helix of PRD2

Phase combination of the MAD and MR phases was applied to complete the structure. The structure is actually refined against native data at 1.9A resolution (collected at ID14-1). The structure shows a huge difference in the quaternary structure between the native and the activated PRD. Considerable relative slide movements are also observed in the helices of the PRD2. This explains why the structure could not be solved by MR. In the native structure, the dimerisation interface has completely changed and the phosphorylable histidines have become accessible to phosphorylating enzymes. This structure will be an enormous progress in the understanding of conditional transcriptional antitermination.

Luc Bousset, Nicolas Fay, Ronald Melki (1,25 shifts): Study of the Ure2p structure and of fibrils formation by the protein

The [URE3] phenotype in yeast *Saccharomyces cerevisiae* is due to an altered prion form of Ure2p, a protein involved in nitrogen catabolism. To understand possible conformational changes at the origin of prion propagation, we previously solved the crystal structure of the Ure2p functional region [Bousset et al. (2001) *Structure 9*, 39-46]. We showed the protein to have a fold similar to that of the, class of glutathione *S*-transferases (GSTs). Here we report crystal structures of the Ure2p functional region (extending from residues 95-354) in complex with glutathione (GSH), the substrate of all GSTs, and two widely used GST inhibitors, namely, *S*-hexylglutathione and *S*-*p*-nitrobenzylglutathione [Bousset et al. (2001) Biochemistry. 40 :13564-7]

Two sets of X-ray diffraction data were collected at 2.5 Å and 2.9 Å resolution from crystals of Ure2p 95-354 soaked in the presence of two concentrations of ADAN-glutathione a potential ligand of the protein. These data sets are of sufficient quality to allow us to calculate electron density maps. However, the ligand does not show in the electron density maps suggesting either a low affinity or an insufficient saturation of the binding sites. The concentration of the ligand will be increased for the next set of data collection.

X-ray fiber diffraction patterns of Ure2p fibrils assembled under different experimental conditions were also collected. These patterns reveal in some cases cross beta sheet structures. This is part of an ongoing project. A paper summarizing previous data collected at the ESRF is submitted.

Sebastiano Pasqualato, Agnidipta Ghosh, Louis Renault (4,25 shifts): Regulation by GTP-binding proteins

We have used 4.25 shifts in 16 bunch on FIP to test diffraction of small and unique crystals and to collect native data sets of different projects related to regulatory GTP-binding proteins.

1) Diffraction Tests on crystals of an inhibited Arf-GDP-ArfGEF complex:

Since GTP-binding proteins have high affinity for nucleotides with very slow intrinsic dissociation rates, they need to interact with Guanine Exchange Factor (GEF) to become instantaneously activated by exchanging GDP for GTP. The GDP→GTP exchange reaction is a complex multi-step reaction which is poorly understood. We work on small GTP-binding proteins Arf involved in the budding of transport vesicles between the Endoplasmic Reticulum and the plasma membrane. To understand the initial steps of the reaction which are constituted by transient ternary **Arf-GDP-ArfGEF** complexes of weak affinity, we had obtained small and unique crystals of such complexes inhibited by a drug. Few small and unique crystals were tested on FIP but did not show any diffraction.

2) The interferon-gamma-induced GTPase GBP1 in its monomeric GMP-bound state:

The Guanylate-Binding Protein 1 belongs to a special class of large GTP-binding proteins of 60-100 kDa whose functions are not yet clear. The protein is induced by interferon gamma which is an immuno-modulatory substance and has an antiviral activity (Anderson S. et al. (1999), Virology 256, 8-14). Its cellular function is related at the biochemical level to the ability to undergo oligomerisation with a high concentration-dependent GTPase activity (Prakash et al. (2000), Nature 403, 567-571). To understand at the molecular level the oligomerisation and high and unique GTPase activity properties of the protein, we have trapped and cristallized the protein in different intermediate states of the GTPase reaction by using non-hydrolysable nucleotide analogues and Aluminium Fluoride.

On FIP we have collected a first complet data set of 3.0-2.8 Å of GBP1 crystallized with GMP. For this data set we have collected 220° with 0.35° oscillations per frames (for 20h) because of an ambiguity to index the data set between different space groups of low symmetry (triclinic to monoclinic) and with an eventual big unit cell parameter of 350 Å. We have finally been able to index the data set based on a complete data set in P1: the crystal belongs to P21 with a unit cell of (a,b,c)=(40Å, 173.5 Å, 49.7 Å) et (α,β,γ)=(90°,110.7°,90°). First electron density maps obtained on the beam line and after initial refinement confirm that the indexing is correct and that GMP is bound to the protein. The structure is currently being refined. It will reveal the GBP1 conformation corresponding to the final monomeric state of the GTP hydrolysis reaction.

3) The small GTP-binding protein Rab11 in its active GTP-bound state:

Rab proteins are small GTP binding proteins involved in tethering and specific sorting between vesicles and membranes, allowing the vesicles to specifically address to the target membranes and fuse with them in order to release their cargo. Rab11a is implicated in the regulation of the trafficking events in the endocytic pathway at the recycling endosome.

As all other small nucleotide binding proteins it cycles between an 'inactive' GDP-bound conformation and an 'active' GTP-bound form that can bind effectors proteins and thus transmits a signal or trigger an event. Given the high number of Rab proteins in the cell (as high as 60 in human), understanding the structural mechanism by which each Rab protein is specifically associated to a given pathway, in which it plays a key role, remains a major issue.

To elucidate the structural GDP/GTP cycle of Rab11a we undertook its crystallization in complex with GDP, the non-hydrolysable analog of GTP, GTPγS, and GTP. The structures of Rab11a in complex with GDP and GTPγS have been solved by virtue of data collected at the ESRF in late 2001 and early 2002. In November 11th, 2002, we have collected a complete dataset in I422, at 1.7 Å of Rab11a(Q70L) in complex with GTP, on beamline BM30. The structure is currently under refinement.