ESRF	**Experiment title: Structural investigation of initiation of DNA replication**	**Experiment number**: 1633
Beamline:	**Date of experiment**: from: 5 May 2000 to: 7 May 2000	**Date of report**: 27 Feb 2001
Shifts:	**Local contact(s)**: Eric Fanchon	*Received at ESRF:*

Names and affiliations of applicants (* indicates experimentalists):

***Dr J.B. Rafferty - University of Sheffield**

Prof. D.W. Rice – University of Sheffield

*Dr P.J. Baker – University of Sheffield

Dr P.J. Artymiuk – University of Sheffield

Report:

The experiment was aimed at the structure determination of the G39P protein, one of the components of the initiation of DNA replication machinery from the *B.subtilis* SPP1 phage system. This work forms part of an ongoing E.C. funded collaboration with the laboratory of Dr J.Alonso in Madrid and involves the structural characterization of all the components and their complexes. In particular, the replication origin binding protein (G38P), the DNA helicase (G40P) and the loader/activator (G39P) of SPP1 and their complexes with each other and DNA targets are being studied. All three protein components have been purified and crystals had been grown of G39P (for which space group and cell parameters had been determined: $P6_122/6_522$ a=b=105Å c=46Å $\alpha=\beta=90°$ $\gamma=120°$). The G39P crystals diffracted with a d_{min} of approximately 3.3Å but numerous heavy atom derivative experiments had been unsuccessful in producing phases for an MIR structure determination. Thus the G39P protein had been sought in a selenomethionine incorporated form. However, after numerous crystallization experiments and analyses of the protein by SDS-PAGE and electrospray mass spectrometry in the time that elapsed between the beam time application and the data collection trip, it became clear that there was a serious problem with proteolytic cleavage which was effecting reproducibility of crystallization and diffraction qulaity. Thus a C-terminal truncated mutant was engineered that proved to be more resistant to proteolysis and this was introduced into an *E.coli* overexpression system. Overexpressed mutant protein (G39Pm) was purified and crystallized but in a different space group to that seen previously for the full length protein (space group $P2_12_12_1$; a= 86Å b= 90Å c= 48Å $\alpha=\beta=\gamma=90°$; 3 copies of the 13.6kDa monomer in the asymmetric unit (a.u.) and a Vm of 2.3). These crystals were grown more reproducibly and the diffraction quality had improved with a d_{min} on home lab. sources of 2.3Å. The crystals could be readily cryo-protected and a selenomethionine incorporated form of the protein was produced. The presence of selenium was checked by electrospray mass spectrometry and the level of incorporation estimated to be about 100%. The protein was crystallized for use in MAD phasing experiments

at the ESRF and these crystals were assessed for diffraction quality in Sheffield before being transported to the ESRF pre-frozen in a “dry-shipper” transport dewar.
At the ESRF, a fluorescence scan of the crystals on station BM30 confirmed the presence of a clear absorbance edge characteristic of selenium around 0.98Å and three wavelengths were selected on the basis of the scan to maximize the differences in the f’ and f” signals in the diffraction measurements from the crystal in a full three wavelength MAD phasing experiment. The selected wavelengths were 0.979473Å (the inflection point of the scan, designated λ^1), 0.979281Å (the fluorescence peak, designated λ^2)and 0.971671Å (a wavelength remote from the peak on the high energy side, designated λ^3). Estimates of the values for the f’ and f’’ components of the diffraction intensities at each wavelength were derived from the fluorescence scans. The data were collected on a MAR345 detector to a d_{min} value of 2.4Å in the order $\lambda^1 \rightarrow \lambda^2 \rightarrow \lambda^3$ using inverse beam geometry. Thus for each wavelength a total of 2x72 single pass oscillation images of 1° width were collected over the same angular range for the same exposure time per image (15s). Thus each data set took approximately 3hrs to collect and the whole experiment including sample preparation and fluorescence scans lasted for approximately 10hrs.
The images were subsequently analysed and diffraction intensities indexed and integrated amplitudes calculated in Sheffield using the DENZO program from the HKL suite. The data were scaled and merged in space group $P2_12_12_1$ using the program SCALEPACK and the following statistics obtained for each of the wavelengths used.

	λ1	λ2	λ3
f’	-13.4	-10.1	-3.0
f”	4.5	8.7	4.2
R_{merge}(I) (%)	4.4 (23.4)	4.4 (23.5)	4.4 (23.4)
Completeness (%)	97.4 (99.5)	97.7 (99.5)	97.6 (99.5)
I/σI >3 (%)	77 (35)	73 (36)	74 (37)
Redundancy >2 (%)	91.4 (90)	91.6 (91)	91.6 (90)

$R_{merge}(I) = (\sum|I_i - \langle I\rangle|/\sum I_i) \times 100$ where I_i is an individual intensity measurement, $\langle I\rangle$ is the mean intensity for that reflection and the summation is over all corresponding measurements. The values in parentheses are those found in the highest resolution shell (2.46-2.40Å).

The positions of the selenium sites were found using the SOLVE program using the full 3 wavelength MAD experimental data option. Of the expected 9 selenium sites (3 per monomer in the a.u.), only 6 could be located and their positions were refined and experimental phases calculated using the program MLPHARE from the CCP4 suite. Electron density maps were calculated at 2.8Å resolution, after solvent flattening with the program DM using the amplitudes from wavelength λ^1 (the inflection point), that corresponded to the two possible solutions for the hand of the heavy atom sites. They showed a clear difference between the two choices and it was possible to identify protein secondary structure features within the map with the correct hand. Subsequent model building was carried using the graphics program TURBO-FRODO to maps that had undergone 3-fold averaging in DM based upon the selenium sites. The resolution of the map was extended out to the limits of the data (2.4Å). When sufficient of the model had been built, it was subjected to maximum likelihood refinement against the experimental data using the program REFMAC. The final model has an R-factor of 21.5% and an R_{free} of 25.1% and comprises 60% of the total number of residues of the structure with the remainder (including the 3 unlocated selenium sites) being disordered within the crystal lattice. In brief, the structure of G39Pm is composed of an N-terminal 3 helical bundle with a disordered C-terminus that is presumed to become ordered upon binding of G39P to one of its partner proteins (G40P or G38P) during the initiation of DNA replication. A manuscript describing the X-ray crystallographic analysis and other biophysical characterizations of G39P is currently in preparation. Attempts at locating the G39Pm structure by molecular replacement within the diffraction data from the full length G39P molecule collected previously have so far been unsuccessful and new efforts are being made to stabilize the full length version of the protein so as to improve crystallization and hence also pursue its structure and those of complexes with G38P and G40P.