Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Crystallization of the *Bacillus subtilis* SPP1 bacteriophage helicase loader protein G*39*P

S. Bailey,[a] S. E. Sedelnikova,[a] P. Mesa,[b] S. Ayora,[b,c] J. C. Alonso[b] and J. B. Rafferty[a]*

[a]Krebs Institute for Biomolecular Research, Department of Molecular Biology and Biotechnology, University of Sheffield, Western Bank, Sheffield S10 2TN, England, [b]Departamento de Biotecnología Microbiana, Centro Nacional de Biotecnologia, CSIC, Campus Universidad Autonoma de Madrid, Cantoblanco, 28049 Madrid, Spain, and [c]Departamento de Biología Molecular, Universidad Autónoma de Madrid, Madrid, Spain

Correspondence e-mail: j.rafferty@sheffield.ac.uk

Received 5 September 2002
Accepted 31 March 2003

The essential helicase loader protein G*39*P encoded by *Bacillus subtilis* SPP1 phage has been overproduced in *Escherichia coli* and purified. The wild-type protein has been crystallized by the hanging-drop vapour-diffusion method in a primitive hexagonal space group, probably $P6_122/P6_522$, but the crystals diffract to only 3.4 Å and are poorly reproducible. Mass-spectrometric analysis has revealed marked proteolytic cleavage from the C-terminus and the presence of a major species corresponding to deletion of the 14 C-terminal residues. Thus, a new variant of the protein (G*39*P112) has been engineered that corresponds to a 14-residue C-terminal truncation. The G*39*P112 variant has also been crystallized but now in a primitive orthorhombic form, probably $P2_12_12$ or $P2_12_12_1$, with unit-cell parameters a = 85.6, b = 89.7, c = 47.6 Å, with diffraction to 2.4 Å on a synchrotron source and with greatly improved reproducibility. Calculation of V_M values for this G*39*P112 variant suggests the presence of three monomers in the asymmetric unit, corresponding to a solvent content of about 47%. A selenomethionine-incorporated form of the protein has been produced and a full three-wavelength MAD data collection undertaken.

1. Introduction

The majority of our understanding of bacterial replicative DNA helicases comes from work with *Escherichia coli* and its extrachromosomal elements (Baker & Bell, 1998) and as such, little is known about Gram-positive replication systems, which are phylogenetically distant from *E. coli*. More recently, work has begun on a Gram-positive system, the *Bacillus subtilis* phage SPP1 replication system. Initiation of SPP1 θ-type DNA replication is strictly dependent on the products of the phage-encoded genes *38*, *39* and *40* (G*38*P, G*39*P and G*40*P, respectively) and the host-encoded DNA primase (DnaG) and DNA polymerase III (Pedré *et al.*, 1994). However, SPP1 replication is independent of the host-encoded replisome organiser (DnaA), the replicative DNA helicase (DnaB-like protein) and molecular chaperones (*e.g.* DnaK, DnaJ or GrpE).

G*40*P, the replicative DNA helicase of SPP1, is a member of the DnaB family of helicases (36% identity to *E. coli* DnaB) and as such is a ring-shaped hexamer that unwinds DNA in the 5′ to 3′ direction (Bárcena *et al.*, 1998; Ayora *et al.*, 2002). G*38*P is an *ori*-type replisome organiser and specifically interacts with the SPP1 replication origin (*ori*L), but does not form a stable interaction with G*40*P (Missich *et al.*, 1997). The helicase loader G*39*P forms a complex with ATP-bound G*40*P and upon binding inactivates the single-stranded DNA-binding, ATPase and unwinding activities of the helicase (Ayora *et al.*, 1999). G*39*P also forms a heterodimeric complex with G*38*P and although both proteins show no significant sequence homology to any other protein, they do share a 15% identity with each other (Pedré *et al.*, 1994; Ayora *et al.*, 1999).

The first step in SPP1 replication initiation involves the ATP-independent binding of multiple copies of G*38*P to the SPP1 replication origin, *ori*L. Once bound, G*38*P is thought to induce local unwinding of an AT-rich sequence located within *ori*L. The G*39*P–G*40*P–ATP complex is then loaded onto the G*38*P-bound origin *via* the interaction between G*39*P and G*38*P (Ayora *et al.*, 1999). This G*39*P–G*38*P interaction results in the dissociation of G*39*P from the G*39*P–G*40*P–ATP complex and the formation of G*39*P–G*38*P heterodimers, thereby relieving the inhibitory effect of the loader on the helicase. Thus, the ATP primed helicase binds the ssDNA at the origin and subsequently hydrolyses ATP during the unwinding of the DNA.

In order to understand the mechanism of inhibition of G*40*P by G*39*P and the interaction of the various component proteins, we have undertaken X-ray crystallographic structural studies. Here, we report the purification and crystallization of both wild-type G*39*P (wtG*39*P) and a C-terminal truncated variant

of G*39*P (G*39*P112), plus the preliminary data from a MAD phasing experiment utilizing a selenomethionine-incorporated form of the variant.

2. Cloning and expression

The cloning and overexpression of the gene encoding wild-type G*39*P has been described elsewhere (Pedré *et al.*, 1994). The mutated SPP1 gene *39* encoding the G*39*P112 variant (*39*K113Amb gene) was cloned into a pT712 vector and the resulting construct, pCB366, was used to transform *E. coli* strain BL21(DE3)[pLysS] and colonies were picked by ampicillin (50 µg ml^{-1}) and chloramphenicol (15 µg ml^{-1}) selection. Overexpression was induced by the addition of 1 m*M* isopropyl thiogalactopyranoside and subsequent addition of rifampicin, as previously described in Ayora *et al.* (1999). G*39*P112 was expressed in a selenomethionine-labelled form (SeMetG*39*P112) using methionine-biosynthesis inhibition by the addition of high concentrations of specific amino acids at the mid-log phase of cell growth (Doublié, 1997, and references therein).

3. Purification

Purification of both wtG*39*P and SeMetG*39*P112 proceeded by a similar protocol. The cell paste was suspended in buffer *A* (40 m*M* Tris–HCl pH 8.0, 2 m*M* EDTA, 10 m*M* dithiothreitol) containing 50 m*M* NaCl and lysed by sonication on ice (four pulses, 20 s each) followed by centrifugation for 15 min at 19 000 rev min^{-1}. The supernatant, which either contained wtG*39*P or SeMetG*39*P112, was loaded onto a 40 ml DEAE-Sepharose Fast Flow column equilibrated with buffer *A* containing 50 m*M* NaCl. wtG*39*P did not bind to the column under these conditions and was collected in the flowthrough fraction; in contrast, SeMetG39P112 did bind and was eluted with a 50–450 m*M* NaCl gradient developed over 300 ml. 4 *M* ammonium sulfate (AS) was then added to the wtG*39*P/SeMetG*39*P112-containing fractions to a final concentration of 1.5 *M*. The pellet was discarded and the supernatant loaded onto a 15 ml phenyl-Toyopearl column pre-equilibrated with buffer *A* plus 1.5 *M* AS. Under these conditions, both wtG*39*P and SeMetG*39*P112 bound to the column and were eluted with a 1.5–0 *M* AS gradient developed over 160 ml. The wtG*39*P/SeMetG*39*P112-containing fractions were then precipitated by addition of solid AS to 40%(*w*/*v*). In the case of wtG*39*P the pellet was resuspended in approximately 5 ml of buffer *B*, equivalent to buffer *A* but at pH 8.5, and dialysed against the same buffer overnight. The dialysate was then loaded on an 8 ml DEAE-Toyopearl column equilibrated with buffer *B* and eluted with a 0–50 m*M* gradient developed over 120 ml. For SeMetG*39*P112, the AS-precipitated pellet was dissolved in buffer *A* plus 100 m*M* NaCl before being loaded onto a Superdex 200 HiLoad gel-filtration column equilibrated with the same buffer and eluted with a 1 ml min^{-1} flow rate. After the last purification step, both wtG*39*P and SeMetG*39*P112 were 99% pure as judged by SDS–PAGE Coomassie blue staining. For crystallization, both proteins were concentrated to 20–30 mg ml^{-1} in buffer *A*. The presence of selenium was assayed in the G*39*P112 protein by electrospray mass spectrometry and the level of incorporation estimated to be greater than 99%.

4. Crystallization and data collection of wtG*39*P

Crystallization of wtG*39*P was achieved at 283 K using the hanging-drop method of vapour diffusion following a preliminary screen using Hampton Research Crystal Screens I and II. Small crystals with hexagonal rod morphology were obtained from condition 3 of screen I. The growth conditions were optimized by increasing the protein concentration to 30–40 mg ml^{-1} and adjusting the precipitant as well as carrying out crystallization at 283 K. Under these modified conditions, crystals were obtained after three weeks using a well solution consisting of 0.2 *M* ammonium phosphate and 0.1 *M* potassium phosphate buffer pH 4.5. Attempts to cryoprotect these crystals for flash-cooling with a range of known cryoprotectants were unsuccessful and therefore the crystals were mounted in capillaries at room temperature for X-ray experiments. Data for wtG*39*P were collected from a single crystal on a MAR 345 image-plate detector mounted on a Rigaku RU-200 X-ray generator with Yale focusing mirror optics.

5. Mass-spectrometric analysis of wtG*39*P

Mass-spectrometric analysis of several wtG*39*P crystals revealed that in addition to G*39*P (molecular weight 14 610 Da), crystals contained several protein species ranging in molecular weight from 13 058 to 14 610 Da. Each of these species could be assigned to the products of proteolytic digestion from the C-terminus of the protein. Furthermore, crystals that diffracted to a higher resolution were found to contain more of a particular species of molecular weight 13 058 Da (corresponding to a G*39*P molecule with 14 C-terminal residues missing) in preference to any other cleavage product (Fig. 1). Subsequent purification and/or crystallization using a range of protease inhibitors in an attempt to completely remove this protease activity failed and indeed inhibited crystallization. Therefore, a truncated variant was engineered, G*39*P112, comprising only the first 112 of the 126 residues of G*39*P.

6. Crystallization and data collection of SeMetG*39*P112

G*39*P112 crystals were obtained after two weeks by sitting-drop vapour diffusion at 293 K, again following a preliminary screen using Hampton Research Crystal Screens I and II but with an additional screen around those conditions that had successfully crystallized wtG*39*P. Small crystals were only obtained from condition 32 of screen I. Optimal crystallization, producing crystals with square-based rod morphology, was found by only varying the precipitant conditions, such that the best crystals were grown using well solutions consisting of a 0.1 *M* sodium citrate buffer pH 4.8–5.2 and 0.8–1.2 *M* AS.

For flash-cooling, a crystal was transferred stepwise through solutions identical to those for optimal crystallization but containing 5,

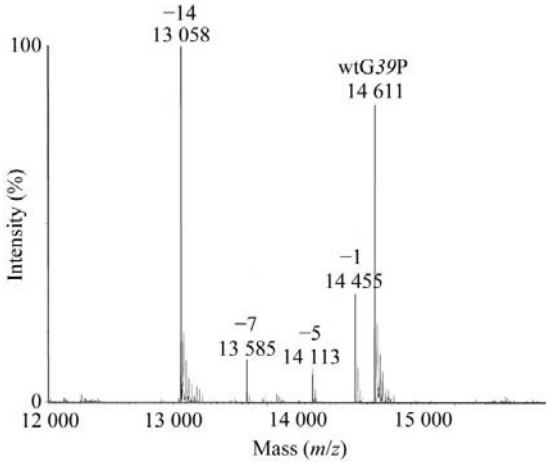

Figure 1
Mass spectra of a wtG*39*P crystal that exhibited diffraction to ∼3.4 Å. Peaks are labelled with the apparent mass as well as either wtG*39*P or the number of residues cleaved from the C-terminus.

Table 1
Data-collection statistics.

Values in parentheses refer to the outer shell.

	wtG*39*P	SeMetG*39*P112 MAD data		
Data set	Native	Inflection	Peak	Remote
Wavelength (Å)	1.5419	0.9795	0.9793	0.9717
Resolution (Å)	20–3.4 (3.48–3.40)	20–2.4 (2.46–2.40)	20–2.4 (2.46–2.40)	20–2.4 (2.46–2.40)
No. of observations	7190	161039	162010	159504
Unique reflections	2243	28352	28373	28331
Data completeness (%)	99.3 (99.0)	93 (81)	97.6 (93.9)	93 (81)
$I/\sigma(I) > 2$ (%)	85.1 (62.0)	77.5 (49.5)	76.6 (49.7)	75.7 (49.9)
$R_{\rm merge}$ (%)	6.6 (32.0)	5.4 (26.2)	5.4 (26.5)	4.4 (30.2)

10, 15, 20 and 25%(*w*/*v*) glycerol before being positioned in a nitrogen-gas stream at 100 K. Data were collected from crystals of the SeMetG*39*P112 form at the European Synchrotron Radiation Facility (ESRF) on station BM30 as part of a three-wavelength MAD phasing experiment.

7. wtG*39*P data analysis

Analysis of the wtG*39*P data using the autoindexing algorithm of the program *DENZO* (Otwinowski & Minor, 1997) showed the crystals have a primitive hexagonal lattice with unit-cell parameters $a = b = 105.3$, $c = 47.4$ Å. Processing and scaling of the data in the *HKL* suite of programs (Otwinowski & Minor, 1997) resulted in an overall $R_{\rm merge}$ of 6.6% and an $R_{\rm merge}$ in the highest resolution shell (3.48–3.4 Å) of 32.0%. Complete data-collection statistics are shown in Table 1. Examination of the diffraction pattern with the program *HKLVIEW* (Collaborative Computational Project, Number 4, 1994) indicates strong 6/*mm* symmetry in all layers and a pattern of systematic absences on the 00*l* axis suggesting the space group to be one of the two enantiomorphs $P6_122$ or $P6_522$. Given a monomer molecular weight of 14 610 Da, packing-density calculations for a monomer in the asymmetric unit give a $V_{\rm M}$ of 2.51 Å^3 Da^{-1}, which corresponds to a solvent content of 51%.

8. SeMetG*39*P112 data analysis

Autoindexing of the SeMetG*39*P112 variant data using the program *DENZO* indicated that the crystals have a primitive orthorhombic lattice with unit-cell parameters $a = 85.6$, $b = 89.7$, $c = 47.6$ Å and therefore a unit-cell volume of $\sim 3.65 \times 10^5$ Å^3. Processing and scaling of the data in the *HKL* suite of programs resulted in the statistics given in Table 1. Inspection of the diffraction pattern using the program *HKLVIEW* revealed that reflections on the *h*00, 0*k*0 axis with *h* and *k* odd, respectively, appeared to be systematically absent, suggesting the space group to be either $P2_12_12$ or $P2_12_12_1$. A self-rotation function was calculated on data in the 10–6 Å resolution range with a 15 Å radius of integration using the program *POLARRFN* (Collaborative Computational Project, Number 4, 1994). The results of this self-rotation function indicated the presence of either threefold or sixfold non-crystallographic symmetry parallel to the crystallographic *c* axis (Fig. 2). $V_{\rm M}$ calculations indicate three copies of the G*39*P112 monomer in the asymmetric unit ($V_{\rm M}$ = 2.33 Å^3 Da^{-1}). There were no significant indications of pseudo-translation revealed by inspection of a native Patterson synthesis. It is envisaged that G*39*P112 has crystallized as either a monomer that packs with a non-crystallographic sixfold axis parallel to the crystallographic *c* axis or a trimer whose threefold axis is parallel to the crystallographic *c* axis, although gel filtration during purification did not suggest the presence of trimers in solution.

Determination of the structure of G*39*P112 by MAD phasing is under way and the selenium substructure is currently being sought. The resulting model of the G*39*P112 variant will be used in an attempt to determine the structure of the wild-type protein by molecular replacement.

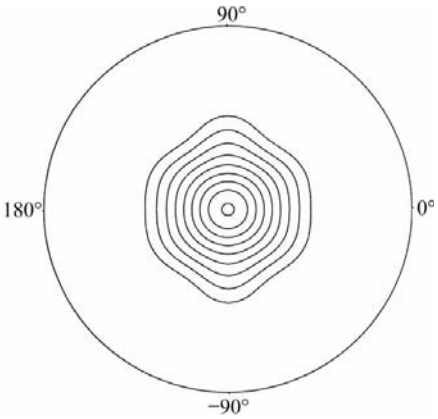

Figure 2
Stereographic projection of the self-rotation function of G39P112 on the $\kappa = 60°$ section.

We thank Michel Roth and the station staff on BM30 at the ESRF. This work was supported by EU grants BIO4-CT98-0106 and QLK2-CT-2000–00634 and the EMBL under the 'Human Capital and Mobility' programme. The Krebs Institute is a designated BBSRC Biomolecular Sciences Centre and a member of the North of England Structural Biology Centre. JBR is a Royal Society Olga Kennard Fellow.

References

Ayora, S., Mesa, P., Weize, F., Stasiak, A. & Alonso, J. C. (2002). *Nucleic Acids Res.* **30**, 2280–2289.
Ayora, S., Stasiak, A. & Alonso, J. C. (1999). *J. Mol. Biol.* **288**, 71–85.
Baker, T. A. & Bell, S. P. (1998). *Cell*, **92**, 295–305.
Bárcena, M., San Martin, C., Weize, F., Ayora, S., Alonso, J. C. & Carazo, J. M. (1998). *J. Mol. Biol.* **283**, 809–819.
Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.
Doublié, S. (1997). *Methods Enzymol.* **276**, 523–530.
Missich, R., Weize, F., Chai, S., Pedré, X., Lurz, R. & Alonso, J. C. (1997). *J. Mol. Biol.* **270**, 50–64.
Otwinowski, Z. & Minor, W. (1997). *Methods. Enzymol.* **276**, 307–326.
Pedré, X., Weize, F., Chai, S., Luder, G. & Alonso, J. C. (1994). *J. Mol. Biol.* **236**, 1324–1340.