Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Overexpression, purification, crystallization and data collection of a single-stranded DNA-binding protein from *Sulfolobus solfataricus*

Iain D. Kerr,† Ross I. M. Wadsworth,† Wulf Blankenfeldt, Adam G. Staines, Malcolm F. White and James H. Naismith*

The Centre for Biomolecular Sciences, The University, St Andrews KY16 9ST, Scotland

† These authors made an equal contribution.

Correspondence e-mail: naismith@st-and.ac.uk

Received 13 June 2001
Accepted 9 July 2001

Single-stranded DNA-binding proteins are recruited when single-stranded DNA is exposed by disruption of the duplex. Many important biological processes such as DNA replication can only occur when the two strands of the duplex are separated. A defining trait of these proteins is the presence of the so-called OB fold. The single-stranded DNA-binding protein of the crenarchaeote *Sulfolobus solfataricus* has a number of interesting differences and similarities to both the eubacterial and eukaryotic homologues. It has an extended C-terminal tail with significant sequence identity to a similar region in the eubacterial protein. However, the sequence of the OB fold is much more like the eukaryotic and euryarchaeal proteins. The *S. solfataricus* protein remains a monomer in the absence of DNA but rapidly polymerizes upon binding – a behaviour not seen in the *Escherichia coli* protein. The protein has been overexpressed, purified and crystallized. The protein crystallizes in two related forms, both having space group $P6_1$ (or $P6_5$) with approximate unit-cell parameters $a = b = 75$, $c = 69$ Å, but the crystals are distinguished by their size and morphology. The larger crystals are hexagonal bipyramids and are merohedrally twinned, diffracting to 1.34 Å with diffraction observed to 1.2 Å. Smaller needle-like crystals diffract to about 2.0 Å but are not twinned. Molecular-replacement attempts have failed owing to low identity with available search models. The structure will be determined by multiple-wavelength methods.

1. Introduction

The disruption of double-stranded DNA can arise as a consequence of strand separation during replication, transcription, recombination and repair. Single-stranded DNA-binding proteins (SSBs) are ubiquitous in nature and play a central role in information-processing pathways by binding single-stranded DNA (ssDNA). In all eukaryotes from yeast to human, the major SSB protein is replication protein A (RPA), a heterotrimeric molecule composed of 70, 32 and 14 kDa subunits (reviewed in Wold, 1997). The two larger subunits each contain a motif known as the 'OB-fold' (oligonucleotide/oligosaccharide/oligopeptide binding fold; Murzin, 1993), a domain of approximately 100 amino acids that can bind roughly three nucleotides of ssDNA. RPA is considered to have four or five OB folds that provide the DNA-binding functionality, together with other domains such as a zinc-finger motif that provides interaction surfaces for DNA replication and repair proteins. The crystal structures of human RPA (Bochkarev *et al.*, 1997, 1999), *E. coli* SSB and human mitochondrial SSB (Webster *et al.*, 1997; Raghunathan *et al.*, 2000; Yang *et al.*, 1997) reveal the universality of the OB fold in this class of proteins, although the eubacterial proteins bring four OB folds together in a homotetrameric rather than heterotrimeric structure.

The third domain of life, the archaea, appear to share significant similarities with eukaryotes in the means by which they replicate, repair and transcribe the genetic material (reviewed in Keeling & Doolittle, 1995). This has resulted in the study of archaeal information-processing pathways as a model for the more complex eukaryotic equivalents. In particular, archaeal proteins have proven amenable to structural studies, as witnessed by the recently published crystal structures of archaeal primase (Augustin *et al.*, 2001), DNA polymerase (Hashimoto *et al.*, 2001), a DEAD-box helicase (Story *et al.*, 2001) and the PCNA sliding clamp (Matsumiya *et al.*, 2001). The archaea are deeply subdivided into the euryarchaeal and crenarchaeal families. The former have essentially eukaryotic-like RPA homologues that can be heterotrimeric (Komori & Ishino, 2001), although they may also exist as monomers and heterodimers (Chedin *et al.*, 1998), often containing the zinc-finger motif found in eukaryotic RPA (Kelly *et al.*, 1998). The

crenarchaea, however, possess SSB homologues that appear to resemble the bacterial SSB proteins more closely. In particular, *S. solfataricus* SSB has some similarity to the eubacterial protein in that both exist as homotetramers with a long flexible C-terminal tail (Wadsworth & White, 2001). The tail of the *E. coli* protein has a role in protein–protein interactions and in the recruitment of repair enzymes to sites of DNA damage (Genschel *et al.*, 2000). Tryptic digestion of *Sulfolobus* SSB results in specific cleavage of the C-terminal tail after Arg119, yielding a truncated protein consisting of a single OB fold that, like the *E. coli* protein, retains the ability to bind ssDNA (Wadsworth & White, 2001).

Despite the similarity of their C-terminal sequences, the *Sulfolobus* OB-fold sequence is quite dissimilar to that in eubacterial sequences and bears a closer resemblance to its euryarchaeal and eukaryotic counterparts. The structures of OB folds have recently been shown to exhibit more variety than previously thought. An OB fold present in human RPA has recently been implicated as having a regulatory function in the protein (Bochkareva *et al.*, 2001). The structure of the RPA70 subunit, responsible for the major ssDNA-binding activity, suggests formation of a pseudo-dimer between the two DNA-binding domains (DBDs), DBD-A and DBD-B. The intermolecular interactions responsible for this event also seem to be involved in the binding of small proteins. In the absence of ssDNA, RPA binds to and inhibits sequence-specific dsDNA-binding activity of the tumour suppressor protein p53 (Miller *et al.*, 1997). The presence of ssDNA abolishes the RPA–p53 interaction, thereby activating p53 for sequence-specific DNA duplex binding. By extending a loop into one of RPA's DBDs, p53 may compete with ssDNA in binding to RPA, thus regulating RPA activity. In addition, *Sulfolobus* SSB also has some unique features, notably that it adopts a monomeric structure in the absence of DNA but can multimerize in an ssDNA-dependent fashion (Wadsworth & White, 2001).

We report here the overexpression, crystallization and native data collection for crystals of SSB. The crystal structure will facilitate the investigation of DNA recombination and repair in crenarchaeotes and may help to decipher the more complex pathways in the eukaryotic system.

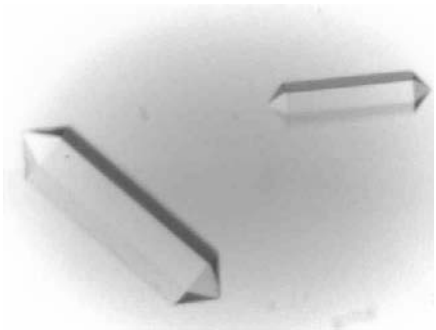

Figure 1
Crystals of SSB.

2. SSB overexpression, purification and modification

Recombinant SSB from *S. solfataricus* has been previously cloned and overexpressed by Wadsworth & White (2001). The intact protein did not give any crystals under Hampton Crystal Screens 1 or 2 (Cudney *et al.*, 1994). To facilitate crystallization the recombinant protein was digested with trypsin, cleaving the last 29 residues from the C-terminus. Analysis of the truncated protein by SDS–PAGE and MALDI-TOF gives a molecular weight of 12 933 Da, corresponding to a single peak in the mass spectrum (Wadsworth & White, 2001).

3. SSB crystallization

Initial conditions were obtained in a Hampton ammonium sulfate grid screen (Jancarik & Kim, 1991; Cudney *et al.*, 1994) at 298 K using the hanging-drop vapour-diffusion method. A protein concentration of 7 mg ml^{-1} and a drop made up of 4 µl protein and 4 µl precipitant was used. Crystals of indistinguishable quality were obtained from a range of ammonium sulfate concentrations in the precipitant (1.4–2.4 *M*). A fairly broad range of 100 m*M* citric acid pH 3.5–5.0 could be used as the buffer. In some wells, needle-like crystals (up to 0.2 × 0.2 × 1.0 mm) were obtained in 2 d, with larger hexagonal bipyramid crystals (0.4 × 0.4 × 0.8 mm) forming in other wells 1–2 weeks later (Fig. 1). However, we have been unable to select for one form by adjusting conditions. Seemingly identical drops can give different results.

4. Data collection

A 1.98 Å data set was collected in-house at 130 K from a single needle crystal mounted on a loop. Crystals were pulled out of the drop through a layer of paraffin oil for cryogenic data collection. Data were collected as 10 min 1° oscillations using a Nonius DIP2000 image plate/rotating anode combination with Osmic mirrors. The crystal-to-detector distance was 94 mm. The data were integrated with *DENZO* and merged with *SCALEPACK* (Otwinowski & Minor, 1996). SSB crystals are primitive hexagonal, with unit-cell parameters $a = b = 75.4$, $c = 69.1$ Å. Systematic absences showed a clear sixfold screw axis identifying the space group as $P6_1$ or $P6_5$. Given the above unit cell, the V_M value (Matthews, 1968) was 2.18 Å^3 Da^{-1}, assuming two molecules per asymmetric unit, with a solvent content of 43%.

A 1.34 Å data set on a bipyramidal crystal was collected at 130 K on Daresbury Laboratory station PX 9.6 using a CCD image plate as the detector. Data were collected in three passes to ensure completeness and high redundancy at all resolutions. The data extend to 1.2 Å, although the quality of the data at this resolution is poor. The data were integrated with *MOSFLM* (Leslie, 1992) and merged with *SCALA* (Evans, 1997). These data have similar unit-cell parameters and the same space group as the other form (Table 1). From an intensity-distribution analysis performed by the *CCP*4 program *TRUNCATE* (Collaborative Computational Project, Number 4, 1994), it was clear that the crystal behaved anomalously. We investigated the possibly of merohedral twinning with the Twinning server (Yeates, 1997; http://www.doe-mbi.ucla.edu/Services/Twinning/) and the *CCP*4 program *DETWIN*. Both programs estimate that the data is approximately 30% twinned. On station PX 9.6 at Daresbury Laboratory, the needle form of the SSB does not diffract to beyond 1.5 Å. We have therefore detwinned the 1.34 Å data using *DETWIN*

Table 1
Data-collection statistics.

a, data collected at Daresbury station PX 9.6; *b*, data collected in-house. Values in parentheses refer to the highest resolution shell.

	a	*b*
Wavelength (Å)	0.87	1.54
Resolution (Å)	47–1.34 (1.39–1.34)	20–1.98 (1.98–1.95)
Space group	$P6_{1/5}$	$P6_{1/5}$
Temperature (K)	130	130
Detector	ADSC CCD	DIP2000
Unit-cell parameters (Å)	$a = b = 75.81$, $c = 70.12$	$a = b = 75.40$, $c = 69.06$
V_M† (Å^3 Da^{-1})	2.25	2.18
Solvent (%)	44.4	43.4
Unique reflections	51077	16159
$I/\sigma(I)$	4.9 (3.1)	32 (6.6)
Average redundancy	6.6 (3.7)	10.7 (5.9)
Data completeness (%)	99.9 (99.9)	96.3 (99.9)
R_{merge}‡ (%)	7.8 (23)	7.1 (26)

† Two molecules per asymmetric unit. ‡ $R_{merge} = \sum\sum I(h)_j - \langle I(h)\rangle / \sum I(h)_j$, where $I(h)$ is the measured diffraction intensity and the summation includes all observations.

and will use these data for refinement when the structure is determined. We have now collected three data sets on each crystal form. In each case the needles show no twinning and in every case the hexagonal bipyramids are merohedrally twinned (30–40%).

5. Molecular replacement

Our attempts at molecular replacement in $P6_1$ and $P6_5$ using either monomer or dimer search models from all the structures mentioned in the introduction have been unsuccessful. Both the merohedrally twinned data and the non-twinned data can, at low resolution (4.5 Å), be merged ($R_{\rm merge} < 10\%$) in the higher symmetry space group $P6_122$ (and $P6_522$). This is consistent with the self-rotation plot, which shows a twofold close to the position expected for a crystallographic twofold. We have used this 'incorrectly' merged data (including the detwinned high-resolution data) for molecular-replacement attempts using a monomer search model without success. We have prepared selenomethionine crystals and confirmed incorporation of selenium by mass spectroscopy. We expect to solve the structure by multiwavelength methods using BM14-UK.

We gratefully acknowledge the use of beamlines at the SRS Daresbury Laboratory. MFW is a Royal Society University Research Fellow and JHN is a BBSRC Research Development Fellow.

References

Augustin, M. A., Huber, R. & Kaiser, J. T. (2001). *Nature Struct. Biol.* **8**, 57–61.
Bochkarev, A., Bochkareva, E., Frappier, L. & Edwards, A. M. (1999). *EMBO J.* **18**, 4498–4504.
Bochkarev, A., Pfuetzner, R. A., Edwards, A. M. & Frappier, L. (1997). *Nature (London)*, **385**, 176–181.
Bochkareva, E., Belegu, V., Korolev, S. & Bochkarev, A. (2001). *EMBO J.* **20**, 612–618.
Chedin, F., Seitz, E. M. & Kowalczykowski, S. C. (1998). *Trends Biochem. Sci.* **23**, 273–277.
Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.
Cudney, R., Patel, S., Weisgraber, K., Newhouse, Y. & McPherson, A. (1994). *Acta Cryst.* D**50**, 414–423.
Evans, P. R (1997). *Jnt CCP4/ESF–EACBM Newsl. Protein Crystallogr.* **33**, 22–24.
Genschel, J., Curth, U. & Urbanke, C. (2000). *J. Biol. Chem.* **381**, 183–192.
Hashimoto, H., Nishioka, M., Fujiwara, S., Takagi, M., Imanaka, T., Inoue, T. & Kai, Y. (2001). *J. Mol. Biol.* **306**, 469–477.
Jancarik, J. & Kim, S.-H. (1991). *J. Appl. Cryst.* **24**, 409–411.
Keeling, P. J. & Doolittle, W. F. (1995). *Proc. Natl Acad. Sci. USA*, **92**, 5761–5764.
Kelly, T. J., Simancek, P. & Brush, G. S. (1998) *Proc. Natl Acad. Sci. USA*, **95**, 14634–14639.
Komori, K. & Ishino, Y. (2001). In the press.
Leslie, A. G. W. (1992). *Jnt CCP4/ESF–EAMCB Newsl. Protein Crystallogr.* **26**.
Matsumiya, S., Ishino, Y. & Morikawa, K. (2001). *Protein Sci.* **10**, 17–23.
Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.
Miller, S. D., Moses, K., Jayaraman, L. & Prives, C. (1997). *Mol. Cell. Biol.* **17**, 2194–2201.
Murzin, A. G. (1993). *EMBO J.* **12**, 861–867.
Otwinowski, Z. & Minor, W. (1996). *Methods Enzymol.* **276**, 307–326.
Raghunathan, S., Kozlov, A. G., Lohman, T. M. & Waksman, G. (2000). *Nature Struct. Biol.* **7**, 648–652.
Story, R. M., Li, H & Abelson, J. N. (2001). *Proc. Natl Acad. Sci. USA*, **98**, 1465–1470.
Wadsworth, R. I. & White, M. F. (2001). *Nucleic Acids Res.* **29**, 914–920.
Webster, G., Genschel, J., Curth, U., Urbanke, C., Kang, C. & Hilgenfeld, R. (1997). *FEBS Lett.* **411**, 313–316.
Wold, M. S. (1997). *Annu. Rev. Biochem.* **66**, 61–92.
Yang, C., Curth, U., Urbanke, C. & Kang, C. (1997). *Nature Struct. Biol.* **4**, 153–157.
Yeates, T. O. (1997). *Methods Enzymol.* **276**, 344–358.