

Cloning, expression, purification, crystallization and preliminary X-ray characterization of the full-length single-stranded DNA-binding protein from the hyperthermophilic bacterium *Aquifex aeolicus*

David J. Clarke,^a Christopher G. Northey,^b Lynsey A. Mack,^b Iain W. McNae,^b Dmitriy Alexeev,^b Lindsay Sawyer^b and Dominic J. Campopiano^{a*}

^aSchool of Chemistry, West Mains Road, University of Edinburgh, Edinburgh EH9 3JJ, Scotland, and ^bStructural Biochemistry Group, Institute of Cell and Molecular Biology, Mayfield Road, University of Edinburgh, Edinburgh EH9 3JR, Scotland

Correspondence e-mail:
dominic.campopiano@ed.ac.uk

Received 14 April 2004
Accepted 18 August 2004

Single-stranded DNA-binding (SSB) proteins stabilize single-stranded DNA, which is exposed by separation of the duplex during DNA replication, recombination and repair. The SSB protein from the hyperthermophile *Aquifex aeolicus* has been overexpressed in *Escherichia coli*, purified and characterized and crystals of the full-length protein (147 amino acids; M_r 17 131.20) have been grown by vapour diffusion from ammonium sulfate pH 7.5 in both the absence and presence of ssDNA [dT(pT)₆₈]. All crystals diffract to around 2.9 Å resolution and those without bound DNA (native) belong to space group $P2_1$, with two tetramers in the asymmetric unit and unit-cell parameters $a = 80.97$, $b = 73.40$, $c = 109.76$ Å, $\beta = 95.11^\circ$. Crystals containing DNA have unit-cell parameters $a = 108.65$, $b = 108.51$, $c = 113.24$ Å and could belong to three closely related space groups ($I222$, $I2_12_12_1$ or $I4_1$) with one tetramer in the asymmetric unit. Electrospray mass spectrometry of the crystals confirmed that the protein was intact. Molecular replacement with a truncated *E. coli* SSB structure has revealed the position of the molecules in the unit cell and refinement of both native and DNA-bound forms is under way.

1. Introduction

Single-stranded DNA-binding (SSB) proteins have been shown to play an essential role in many aspects of DNA metabolism (Chase & Williams, 1986). They preferentially bind and protect vulnerable single-stranded DNA (ssDNA), which is formed transiently during DNA replication, recombination and repair. SSB proteins are characterized by the presence of a conserved OB-fold motif (oligonucleotide/oligosaccharide/oligopeptide-binding fold), which is typically 100 amino acids in length (Murzin, 1993).

SSB proteins can be divided into two distinct groups based on their quaternary structure. Eukaryotic SSB proteins, known as replication protein A (RPA), are exemplified by human RPA, which has a heterotrimeric structure comprising three subunits: RPA70, RPA32 and RPA14 (of molecular weights 70, 32 and 14 kDa, respectively; Wold, 1997). The RPA complex contains six OB folds, four of which bind DNA: three on RPA70 and one on RPA32 (Bastin-Shanower & Brill, 2001). An N-terminal domain on RPA70 has also been shown to be involved in protein–protein interactions (Jacobs *et al.*, 1999). In contrast, bacterial SSB proteins form homotetramers, with each subunit containing one DNA-binding domain (Raghunathan *et al.*, 1997). These DNA-binding domains are located at the N-termini of the individual SSB protein subunits and

form the characteristic OB folds. While the N-terminus of each subunit binds ssDNA and contains the homotetramer interface, it is thought that the C-terminal domain is involved in interactions with other protein components of DNA metabolism. The C-terminal domain of bacterial SSB proteins exhibits low sequence homology across species, with the exception of the terminal six residues, which form a highly conserved negatively charged DDDIPF motif. This motif is essential for the function of *Escherichia coli* SSB protein *in vivo* (Curth *et al.*, 1996) and has been shown to interact directly with the 3'–5' ssDNA-degrading exonuclease I (Genschel *et al.*, 2000). The tails of both the *E. coli* and the *Sulfolobus solfataricus* SSB proteins are not involved in DNA binding, but are thought to play roles in mediating protein–protein interactions with other subunits within the DNA polymerase complex (Bruck *et al.*, 2002). There is also evidence that a mutually exclusive interaction between the C-terminal domain of *E. coli* SSB protein, DNA polymerase and primase is utilized as a three-point switch to initiate the exchange of places of these two proteins on DNA (Yuzhakov *et al.*, 1999). Furthermore, a recent report suggests that the interaction between the DNA polymerase and SSB from RB69 (a T4-like bacteriophage) results in an increase in the overall affinity of the SSB protein for ssDNA (Sun & Shamoo, 2003). Finally, Gulbis and coworkers have recently

theoretical weight of full-length *A. aeolicus* SSB protein (17 131.2 Da).

3. Crystallization

Initial crystals were obtained using Molecular Dimensions Structure Screens 1 and 2 and the sitting-drop vapour-diffusion method at 290 K. The drop consisted of 5 μ l protein solution (7 mg ml⁻¹ in buffer C) and 5 μ l precipitant. Over two weeks, small crystals of native protein were observed under three different conditions, with the best quality obtained using 100 mM HEPES pH 7.5, 2% (v/v) PEG 400, 2.0 M (NH₄)₂SO₄ pH 7.5 as the precipitant. After refining the crystallization conditions, larger crystals were obtained after four weeks using the same precipitant at a pH of 7.0. Co-crystallization of the DNA-bound protein was achieved by mixing 7.5 mg ml⁻¹ protein in a 1:1 molar ratio (tetramer:ssDNA) with 69-mer dT(pT)₆₈ (MWG Biotech). The complex was incubated on ice for 60 min and centrifuged (10 min, 35 000g) prior to crystallizations being set up. Each crystallization drop comprised 1.5 μ l protein in 50 mM Tris pH 7.0, 0.1 M NaCl and 1.5 μ l precipitant. All were set up at 290 K. Crystals grew within one week; the best quality crystals were obtained using 100 mM HEPES pH 7.5, 2.3 M (NH₄)₂SO₄.

4. Data collection and processing

Crystals of native protein of approximate dimensions 0.1 \times 0.2 \times 0.2 mm were flash-cooled in a 20% glycerol well solution and X-ray data for the native SSB were collected at 100 K (Cryostream cooler; Oxford Cryosystems, Oxford, England) on a MAR Research 345 imaging plate mounted on an Enraf-Nonius FR591 rotating-anode generator, $\lambda = 1.5418$ Å, fitted with Osmic mirrors and operating at 40 kV, 110 mA. Crystals of similar dimensions were obtained for the DNA-bound form and data were also collected at 100 K on station 14.2 ($\lambda = 0.978$ Å) at the SRS, CLRC Daresbury Laboratory. Analysis of the diffraction data for both crystals using *MOSFLM/SCALA* (Leslie, 1992; Collaborative Computational Project Number 4, 1994) produced the data shown in Table 1 and allowed the assignment of the native crystals to space group *P*₂₁. For the DNA-bound data, similar processing statistics were obtained with space groups *I*222, *I*₂₁₂₁ and *I*₄₁.

For the native SSB data set, a model of the SSB from *Escherichia coli* (PDB code 1qvc; Matsumoto *et al.*, 2000) was used to search for an initial solution using

MOLREP (Vagin & Teplyakov, 1997). The search molecule was trimmed of its flexible loops and amino-acid side chains to produce a tetrameric polyalanine structure. The top rotation-function solution produced a satisfactory translation-function solution that was then used to aid location of the second tetramer. No solution was obtained using the *S. solfataricus* structure as a search model. For the DNA-bound SSB, a multi-copy search with *MOLREP* using the partially refined SSB *Aae* dimer ($R = 0.272$, $R_{\text{free}} = 0.308$) provided solutions, the best of which contained two dimers per asymmetric

unit in each of the three space groups. 20 cycles of rigid-body refinement were followed by ten cycles of restrained refinement. The statistics for this process are also shown in Table 1. Refinement of both crystal forms is currently in progress while attempts are being made to improve the diffraction quality of the crystals.

To ensure that no autolysis of the protein had occurred, a single crystal of DNA-bound SSB was dissolved in 10 mM HEPES pH 8.1 and analysed by SDS-PAGE and ESI-MS, which revealed no significant degradation of the protein. SDS-PAGE analysis produced a

Table 1
Data-collection and reduction statistics.

Values in parentheses are for the highest resolution shell.

Protein	Native	DNA-bound		
Data collection	Home source	SRS Daresbury		
Oscillation range (°)	180 \times 1	120 \times 1		
Space group	<i>P</i> ₂ ₁	<i>I</i> ₄ ₁	<i>I</i> 222	<i>I</i> ₂ ₁ ₂ ₁
Unit-cell parameters				
<i>a</i> (Å)	80.97	108.36	108.65	
<i>b</i> (Å)	73.40	108.36	108.51	
<i>c</i> (Å)	109.76	113.20	113.24	
β (Å)	95.11	90.0	90.0	
Tetramers per AU	2	1	1	1
V_M (Å ³ Da ⁻¹)	2.37	1.86†	1.87†	1.87†
Resolution limits	70–2.92 (3.08–2.92)	78.28–2.80 (2.95–2.80)	79.06–2.80 (2.95–2.80)	
No. observations	100835 (12811)	79739 (676)	79012 (11550)	
No. unique reflections	27790 (3680)	15953 (2324)	16402 (2368)	
$I/\sigma(I)$	9.2 (3.5)	7.5 (1.4)	7.9 (1.4)	
Completeness (%)	98.9 (98.9)	99.9 (99.9)	99.7 (99.7)	
$R_{\text{merge}}^{\ddagger}$	0.057 (0.212)	0.056 (0.525)	0.053 (0.541)	
<i>MOLREP</i> <i>R</i> factor	0.567	0.560	0.520	0.517
<i>MOLREP</i> correlation coefficient	0.359	0.316	0.471	0.471
Restrained refinement <i>R</i> factor	0.425	0.401	0.396	0.400
Restrained refinement R_{free}	0.487	0.507	0.471	0.488

† For one SSB tetramer ($M_r = 68\,400$) and one DNA 69-mer ($M_r = 20\,927$). ‡ $R_{\text{merge}} = \sum_h \sum_i |I_i(h) - \langle I_i(h) \rangle| / \sum_h \sum_i I_{hi}$, where $\langle I_i(h) \rangle$ is the mean intensity of the *i* symmetry-equivalent reflections

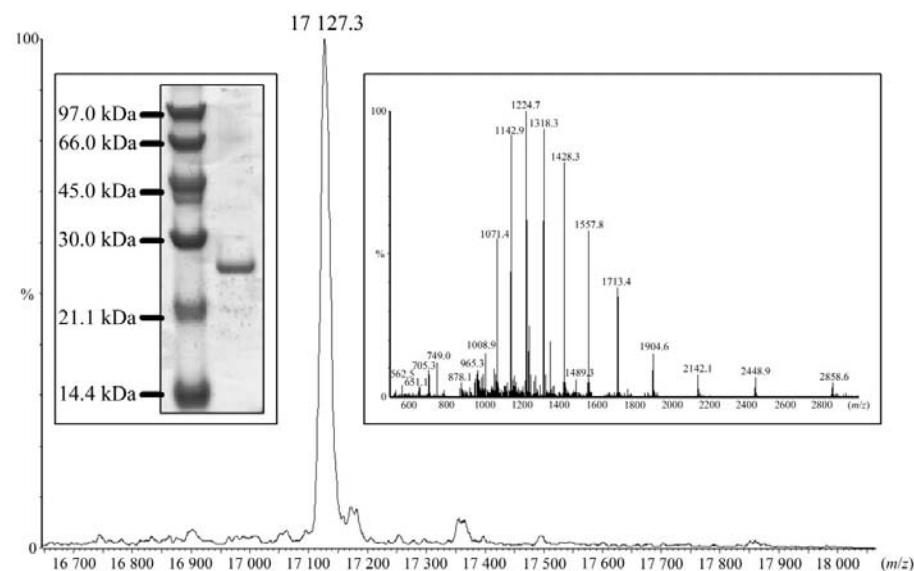


Figure 2
Analysis of the crystallized DNA-bound SSB protein by ESI-MS. The main figure shows the deconvoluted mass of 17 127.3 (obtained using *Transform* software; Micromass UK), consistent with the theoretical value of 17 131.2; right inset, ion envelope of crystallized SSB protein; left inset, SDS-PAGE analysis of crystallized SSB protein results in a single band running at an anomalous weight of \sim 23 kDa.

single band around 23 kDa in keeping with the observed anomalous mobility of the native SSB. Only one major species was observed by ESI-MS with a mass of $17\,127.3 \pm 2.7$ (Fig. 2), in good agreement with the predicted weight. No LCMS data could be obtained from the dissolved native SSB crystal.

5. Discussion

It is clear from the sequence alignment of the SSB proteins that the *A. aeolicus* and the *E. coli* proteins are more closely related to each other than either is to *S. solfataricus* SSB (Fig. 1). This is borne out by the fact that a molecular-replacement solution using the *E. coli* structure was obtained relatively easily, whilst no satisfactory solution could be obtained with the *S. solfataricus* structure. The very high resolution of *S. solfataricus* SSB reveals why this should be so in that the actual molecular structure is much more closely related to the eukaryotic SSB fold than that of *E. coli* SSB (Kerr *et al.*, 2003). Consequently, despite a modest sequence identity, the structures are distinct.

For the DNA-bound crystals reported here, there is an ambiguity as regards the space group. Given an SSB tetramer in the asymmetric unit, using the monomer M_r of 17 100 and that of the DNA as 20 927, the V_M can be calculated to be 2.43, 2.11, 1.86 or $1.51 \text{ \AA}^3 \text{ Da}^{-1}$ for zero, 0.5, one or two bound DNA 69-mers per tetramer. The expected 1:1 complex requiring one DNA oligomer per asymmetric unit corresponds to a V_M of $1.86 \text{ \AA}^3 \text{ Da}^{-1}$, which is within the range found by Matthews (1968), albeit quite close to the lower limit. The physiological tetramer as observed in the native structure sits on a crystallographic dyad in both $I4_1$ and $I2_12_1$, whereas the tetramer sits on a screw dyad axis in $I222$. It is impossible for there to be exact twofold symmetry for the SSB tetramer with a single DNA oligomer bound, although a pseudo-twofold arrangement is possible. Given the limited resolution of the present X-ray data and the current state of the refinement, such a situation cannot yet be ruled out. However, if the DNA oligomer is shared between two tetramers in some fashion, this could permit

the DNA-bound tetramer to lie upon a crystallographic dyad, while maintaining four subunits and a single DNA molecule in the asymmetric unit (Ferrari *et al.*, 1994). The initial electron-density maps in each of the three space groups all show extra electron density near regions of the protein expected to bind DNA (Raghunathan *et al.*, 2000). Examination of the maps together with the statistics shown in Table 1 leads us to prefer $I222$ as the space group, but we are continuing to refine all three possibilities. These refinements should clarify this uncertainty and also allow us to estimate the occupancy of the DNA.

In summary, our expression and purification strategy has produced full-length SSB from the hyperthermophile *A. aeolicus* with no autolysis observed by mass spectrometry and SDS-PAGE. The flexible C-terminal tail is present in the crystals reported here, unlike the truncated SSB used in both the *E. coli* and *S. solfataricus* structure determinations. Our initial refinement of the structures of both forms has allowed the clear assignment of the electron density to residues 1–38 and 41–108 and we are currently refining the models in an effort to distinguish the C-terminal residues.

Note added in proof: During the processing of this manuscript a report has been published describing the crystallization of full-length *E. coli* SSB (Savvides *et al.*, 2004).

We wish to thank Professors Karl Stetter and Robert Huber (University of Regensburg) for the kind gift of *A. aeolicus* chromosomal DNA and Professor Jim Naismith and Dr Iain Kerr (University of St Andrews) for kindly providing the coordinates of the SSB fragment from *S. solfataricus* prior to publication. The Biotechnology and Biological Sciences Research Council UK and the University of Edinburgh supported this work (DJC, LAM).

References

Bastin-Shanower, S. A. & Brill, S. J. (2001). *J. Biol. Chem.* **276**, 36446–36453.
 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.
 Bruck, I., Yuzhakov, A., Yurieva, O., Jeruzalmi, D., Skangalis, M., Kuriyan, J. & O'Donnell, M. (2002). *J. Biol. Chem.* **277**, 17334–17348.
 Chase, J. W. & Williams, K. R. (1986). *Annu. Rev. Biochem.* **55**, 103–106.
 Collaborative Computational Project, Number 4 (1994). *Acta Cryst. D* **50**, 760–763.
 Curth, U., Genschel, J., Urbanke, C. & Greipel, J. (1996). *Nucleic Acids Res.* **24**, 2706–2711.
 Deckert, G., Warren, P. V., Gaasterland, T., Young, W. G., Lenox, A. L., Graham, D. E., Overbeek, R., Snead, M. A., Keller, M., Aujay, M., Huber, R., Feldman, R. A., Short, J. M., Olsen, G. J. & Swanson, R. V. (1998). *Nature (London)*, **392**, 353–358.
 Ferrari, M. E., Bujalowski, W. & Lohman, T. M. (1994). *J. Mol. Biol.* **236**, 106–123.
 Genschel, J., Curth, U. & Urbanke, C. (2000). *J. Biol. Chem.* **381**, 183–192.
 Gulbis, J. M., Kazmirski, S. L., Finkelstein, J., Kelman, Z., O'Donnell, M. & Kuriyan, J. (2004). *Eur. J. Biochem.* **271**, 439–449.
 Jacobs, D. M., Lipton, A. S., Isern, N. G., Daughdrill, G. W., Lowry, D. F., Gomes, X. & Wold, M. S. (1999). *J. Biomol. NMR*, **14**, 321–331.
 Kerr, I. D., Wadsworth, R. I. M., Cubeddu, L., Blankenfeldt, W., Naismith, J. H. & White, M. F. (2003). *EMBO J.* **22**, 2561–2570.
 Leslie, A. G. W. (1992). *Int CCP4/ESF-EACMB Newsl. Protein Crystallogr.* **26**.
 Lohman, T. M. & Ferrari, M. E. (1994). *Annu. Rev. Biochem.* **63**, 527–570.
 Matsumoto, T., Morimoto, Y., Shibata, N., Kinobuchi, T., Shimamoto, N., Tsukihara, T. & Yasuoka, N. (2000). *J. Biochem.* **127**, 329–335.
 Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.
 Murzin, A. G. (1993). *EMBO J.* **12**, 861–867.
 Raghunathan, S., Kozlov, A. G., Lohman, T. M. & Waksman, G. (2000). *Nature Struct. Biol.* **7**, 648–652.
 Raghunathan, S., Ricard, C. S., Lohman, T. M. & Waksman, G. (1997). *Proc. Natl Acad. Sci. USA*, **94**, 6652–6657.
 Savvides, S. N., Raghunathan, S., Futterer, K., Kozlov, A. G., Lohman, T. M. & Waksman, G. (2004). *Protein Sci.* **13**, 1942–1947.
 Sun, S. & Shamoo, Y. (2003). *J. Biol. Chem.* **278**, 3876–3881.
 Vagin, A. & Teplyakov, A. (1997). *J. Appl. Cryst.* **30**, 1022–1025.
 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.
 Yuzhakov, A., Kelman, Z. & O'Donnell, M. (1999). *Cell*, **96**, 153–163.