Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Crystallization and preliminary X-ray studies of the single-stranded DNA-binding protein from *Mycobacterium tuberculosis*

K. Saikrishnan,[a] J. Jeyakanthan,[a] J. Venkatesh,[b] N. Acharya,[b] K. Purnapatre,[b] K. Sekar,[c] U. Varshney[b] and M. Vijayan[a]*

[a]Molecular Biophysics Unit, Indian Institute of Science, Bangalore 560 012, India, [b]Department of Microbiology and Cell Biology, Indian Institute of Science, Bangalore 560 012, India, and [c]Bioinformatics Centre, Indian Institute of Science, Bangalore 560 012, India

Correspondence e-mail: mv@mbu.iisc.ernet.in

Received 23 October 2001
Accepted 21 November 2001

Single-stranded DNA-binding proteins play an important role in DNA replication, repair and recombination. The protein from *Mycobacterium tuberculosis* (*Mt*SSB) is a tetramer with 164 amino-acid residues in each subunit. The protein readily crystallizes in space group $P3_121$ (or $P3_221$) at pH 7.4 under appropriate conditions. Under different conditions, but at the same pH, orthorhombic crystals belonging to space group $I222$ or $I2_12_12_1$ were obtained after several months. Similar orthorhombic crystals were obtained when protein samples stored for several months were used for crystallization. The orthorhombic crystals obtained in different experiments, though similar to one another, exhibited variations in unit-cell parameters, presumably on account of different extents of proteolytic cleavage of the C-terminal region. Molecular-replacement calculations using different search models did not yield the structure. As part of attempts to solve the structure using isomorphous replacement, a good mercury derivative of the trigonal crystal has been prepared.

1. Introduction

DNA metabolic processes such as replication, repair, recombination and transcription involve the unwinding of double-stranded DNA (dsDNA), forming transient single-stranded DNA (ssDNA). ssDNAs in the cell are susceptible to attacks by chemically reactive groups and nucleases and are prone to form secondary structures. Nature has evolved a class of protein, the single-stranded DNA-binding protein (SSB), which binds to ssDNA in a sequence-independent manner. By coating the exposed ssDNA, these proteins protect the chemical integrity of the strand from chemical and nuclease attacks. SSBs are ubiquitous and have been identified in viruses, eubacteria, archaea and eukaryotes (Chase & Williams, 1986; Lohman & Ferrari, 1994). Though SSBs from the various lifeforms have very similar functions, they share very little sequence similarity. They display very different DNA-binding capabilities and exist in various oligomeric states.

M. tuberculosis encodes a 164-residue SSB protein (*Mt*SSB) which was cloned, over-expressed and purified by Purnapatre & Varshney (1999). *Mt*SSB shares similar biochemical properties with the extensively studied SSB protein from *Escherichia coli* (*Ec*SSB). Gel filtration and electrophoretic studies show that, like *Ec*SSB, it exists as a homotetramer in solution and requires a minimum of 15–17 nucleotide ssDNA for efficient complex formation (Purnapatre & Varshney, 1999). The crystal structures of *Ec*SSB in both its free state as well as in complex with ssDNA are available (Raghunathan *et al.*, 1997, 2000; Webster *et al.*, 1997; Matsumoto *et al.*, 2000). *Ec*SSB is structurally similar to human mitochondrial SSB (Yang *et al.*, 1997; Webster *et al.*, 1997). The structure of *Ec*SSB can be divided into two domains. The N-terminal domain contains a motif called the oligonucleotide-binding fold (OB-fold) made up of the first 110 residues, while the C-terminal domain is a long stretch of residues with no regular secondary structure (Matsumoto *et al.*, 2000). The crystal structures of many other SSBs, such as human RPA (Bochkarev *et al.*, 1997, 1999), gene 32 protein of bacteriophage T4 (Shamoo *et al.*, 1995) and gene 2.5 protein of bacteriophage T7 (Hollis *et al.*, 2001), reveal the universality of the OB-fold in this class of proteins (Suck, 1997). Recently, the crystallization of SSB from *Sulfolobus solfataricus*, a crenarchaea, has been reported and is predicted to contain the OB-fold (Kerr *et al.*, 2001).

The first 110 residues of *Mt*SSB corresponding to the OB-fold are involved in DNA binding. The C-terminal domain of *Mt*SSB, like that in *Ec*SSB, is rich in glycine and proline and has a stretch of acidic residues at the end. The tail region mediates protein–protein interactions and recruits repair enzymes at the site of DNA damage (Genschel *et al.*, 2000; Handa *et al.*, 2001) and is essential for the *in vivo* functioning of *Ec*SSB (Curth *et al.*, 1996). This region is susceptible to proteolytic cleavage.

Deletion of this region results in a protein that binds to ssDNA with a higher affinity *in vitro*. Despite these similarities, *Mt*SSB fails to form cross-species heterotetramers with *Ec*SSB (Purnapatre & Varshney, 1999), in contrast to many other bacterial SSBs (de Vries *et al.*, 1994). This indicates that the residues involved in stabilization of the tetramers could be different in the two proteins. Sequence comparison indicates that the patch of aromatic and positively charged residues in *Ec*SSB involved in DNA binding are present in *Mt*SSB, although there are differences in detail (Purnapatre & Varshney, 1999).

2. Materials and methods

2.1. Crystallization and preparation of a heavy-atom derivative

Recombinant SSB from *M. tuberculosis* was overexpressed and purified by methods described previously (Purnapatre & Varshney, 1999). The purified protein was stored at 253 K in 250 µl aliquots at a concentration of 5–10 mg ml^{-1} containing 0.5 *M* NaCl, 20 m*M* Tris–HCl pH 7.4 without any protease inhibitor in the medium. The protein was crystallized at 298 K using the hanging-drop method. In all crystallization experiments, the drop was made up of 4 µl of a 5–10 mg ml^{-1} solution of protein in 20 m*M* Tris–HCl buffer pH 7.4 and 1 µl of precipitant. Trigonal crystals (Fig. 1*a*), hereafter referred to as form I, grew when 1 *M* sodium acetate, 500 m*M* sodium chloride and 50–100 m*M* zinc sulfate in 20 m*M* Tris–HCl buffer pH 7.4 was used as precipitant. The same crystal resulted when zinc sulfate was replaced by cadmium sulfate. The crystals grew in 1–2 weeks. Orthorhombic crystals (form II; Fig. 1*b*) grew in 3–5 months when 2 *M* magnesium chloride in the same buffer was used as the precipitant. Similar crystals, but not the trigonal ones, were obtained when the protein sample stored for several months was used for crystallization. In some instances, macroseeding was resorted to in order to grow diffraction-quality crystals. The unit-cell parameters exhibited considerable variation depending on the age of the sample. A heavy-atom derivative of form I was prepared by soaking the crystals in the mother liquor containing 7 m*M* $HgBr_2$ for 52 h.

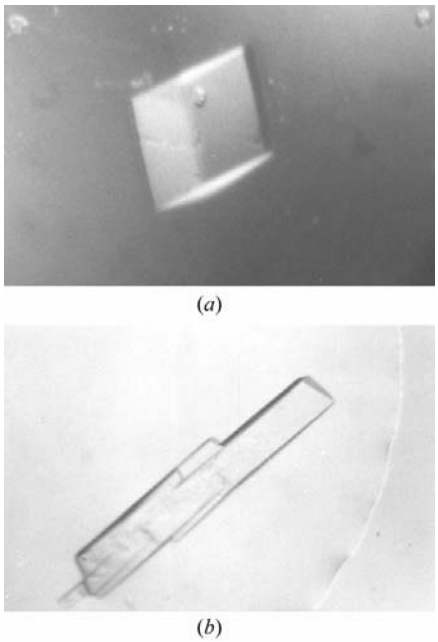

Figure 1
Crystals of *Mt*SSB. (*a*) Form I, (*b*) form II.

Diffraction data from both crystal forms were collected at low temperature (100 K) using flash-freezing with a MAR Research imaging plate mounted on a Rigaku RU-200 X-ray generator. Crystal-to-image plate distance was 120 mm. Data collected at room temperature (298 K) turned out to be of poor quality, especially in the case of form I. These crystals were soaked in glycerol for cryoprotection prior to data collection. The concentration of glycerol was raised stepwise from 0 to 30% over a period of 6 h. Use of this procedure eliminated crystal cracking, reduced mosaicity and improved resolution. No prior processing of crystal was required for form II. The data were processed and scaled using the *HKL* package (Otwinowski & Minor, 1997).

3. Results and discussion

3.1. Heterogeneity in crystals

*Mt*SSB, like *Ec*SSB, has been observed to undergo proteolytic modification upon storage. Proteolytic cleavage has been observed in *Ec*SSB (Ollis *et al.*, 1983; Ng & McPherson, 1989; Matsumoto *et al.*, 2000), but the reason for it is not yet entirely clear, although it has been established that it takes place in the C-terminal region. Recent studies on *Mt*SSB also have indicated that this region is susceptible to proteolytic degradation (Reddy *et al.*, 2001). Freshly purified protein gives a single band corresponding to the native protein (17 350 Da) on SDS–PAGE. Upon storage, minor bands corresponding to truncated protein appear on the PAGE and become prominent over time. Crystals used for data collection were care-

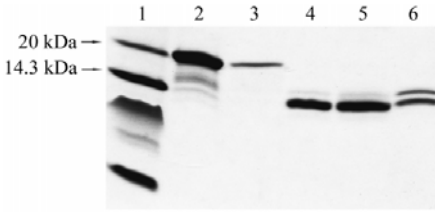

Figure 2
15% SDS–PAGE of purified *Mt*SSB used for crystallization and that recovered from crystals. Proteins were stained with Coomassie Brilliant Blue R250. Lane 1, marker; lane 2, protein sample used for crystallization; lane 3, crystals of form I grown from sample in lane 2; lane 4, protein sample stored for over 12 months; lane 5, crystals of form II obtained from sample in lane 4; lane 6, 130 residues long C-terminal deletion mutant of *Mt*SSB (14 100 Da) and its degradation product.

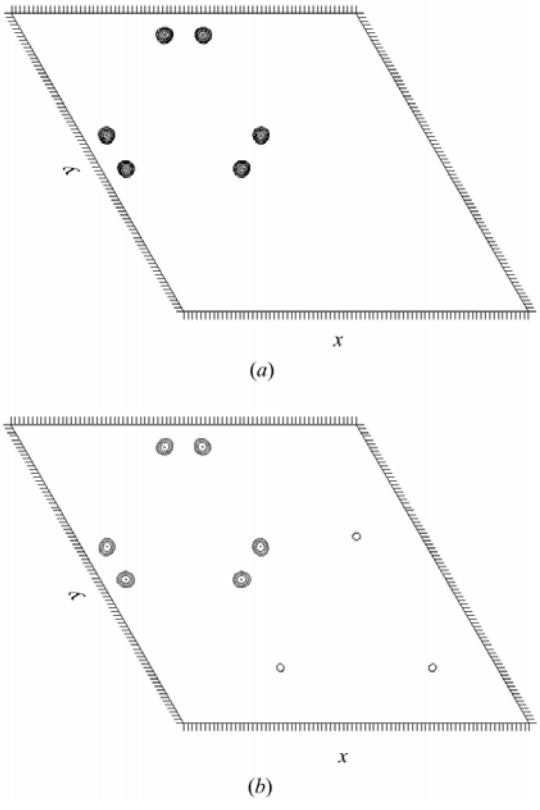

Figure 3
Harker section (z = 1/3) of the difference Patterson maps computed using 10–5 Å resolution data from the native crystal and the mercury derivative crystal. (*a*) Isomorphous difference map. The map is contoured from a start value of 3σ with an interval of 0.5σ. (*b*) Anomalous difference map. The contouring starts at 3σ with an interval of 1σ. The maps were generated using *CNS* (Brunger *et al.*, 1998).

Table 1
Data-collection statistics for form I crystal.

Values in parentheses correspond to the highest resolution shell (3.0–2.9 Å). When Bijvoet pairs were treated separately, 5261 such pairs were obtained, with 568 in the last resolution shell.

	Native	Hg derivative
Resolution (Å)	20.0–2.9	20.0–2.9
No. of observations	53782	55885
No. of unique reflections	6404 (616)	6301 (633)
Compleness (%)	99.9 (99.7)	97.0 (100.0)
$R_{\rm merge}$† (%)	10.8 (67.7)	10.4 (45.1)
$R_{\rm iso}$‡ (%)		25.2 (33.5)
$R_{\rm anom}$§ (%)		4.9 (13.8)
$I/\sigma(I)$	8.7	10.2

† $R_{\rm merge} = \sum\sum_i I_i - \langle I\rangle|/\sum\sum_i I_i$. ‡ $R_{\rm iso} = \sum|I_{PH} - I_P|/\sum I_P$. § $R_{\rm anom} = \sum|I(+) - I(-)|/\sum[I(+) + I(-)]$.

fully washed with the mother liquor, dissolved in double distilled water and then analyzed on SDS–PAGE (Fig. 2). The protein in form I appeared as a single major band corresponding to the intact protein. The available crystal structures of *Ec*SSB are either that of homotetrameric DNA-binding domain of proteolytically treated SSB (Raghunathan *et al.*, 1997, 2000) or heterotetrameric SSB composed of a mixture of native and C-terminal truncated polypeptide chains (Webster *et al.*, 1997; Matsumoto *et al.*, 2000). Form I of *Mt*SSB could be the first bacterial SSB protein crystal with a full-length polypeptide chain. Crystals belonging to form II typically gave two bands, corresponding to approximately 14 and 11 kDa, with varying strength.

3.2. Preliminary X-ray studies

The trigonal crystals (form I) belong to space group $P3_121$ or $P3_221$, with unit-cell parameters $a = b = 78.6$, $c = 77.4$ Å in the case of crystals grown in the presence of zinc sulfate. Crystals with similar unit-cell parameters result when zinc is replaced by cadmium. The crystals diffract to better than 3 Å resolution. The Matthews coefficient (Matthews, 1968) indicates that the crystal contains a dimer in the asymmetric unit, with a solvent content of 38%. The tetramer should therefore be on a crystallographic twofold axis. The orthorhombic crystals belong to space group $I222$ or $I2_12_12_1$. The unit-cell parameters exhibit considerable variability. The best crystals, which diffracted to a resolution of 2.7 Å, have unit-cell parameters $a = 60.4$, $b = 117.6$, $c = 175.2$ Å. The unit cell contains one crystallographically independent tetramer with a solvent content of 55–65% depending on the stoichiochemistry of the 14 and 11 kDa subunits.

Repeated attempts to solve the structure using the molecular-replacement method employing different search models did not succeed, presumably owing to the low sequence identity between them and *Mt*SSB. Among the proteins of known structure, *Ec*SSB is closest to *Mt*SSB in terms of sequence similarity. Even between these two proteins, the sequence identity is as low as 28% (35% in the OB-fold and 13% in the rest of the protein; Purnapatre & Varshney, 1999). A search for heavy-atom derivatives of both forms has so far resulted in a good mercury derivative of the trigonal crystals grown in the presence zinc sulfate. Statistical information on the data from the native and the derivative crystals are given in Table 1. Anomalous differences were clearly discernible in the derivative data. Harker sections of the isomorphous and anomalous difference Patterson maps calculated using the data sets (Fig. 3) indicate the presence of primarily a single heavy-atom position in the derivative. Further analysis of the structure is in progress.

The intensity data were collected at the X-ray Facility for Structural Biology at the Institute, supported by the Department of Science and Technology (DST) and the Department of Biotechnology (DBT). Facilities at the Supercomputer Education and Research Centre, and the Interactive Graphics based facility and Distributed Information Centre (both supported by DBT) were used.

References

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.
Brunger, A. T., Adams, P. D., Clore, G. M., Delano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J.-S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T. & Warren, G. L. (1998). *Acta Cryst.* D**54**, 905–921.
Chase, J. W. & Williams, K., R. (1986). *Annu. Rev. Biochem.* **55**, 103–136.
Curth, U., Genschel, J., Urbanke, C. & Greipel, J. (1996). *Nucleic Acids Res.* **24**, 2706–2711.
Genschel, J., Curth, U. & Urbanke, C. (2000). *Biol. Chem.* **381**, 183–192.
Handa, P., Acharya, N. & Varshney, U. (2001). *J. Biol. Chem.* **276**, 16992–16997.
Hollis, T., Stattel, J. M., Walther, D. S., Richardson, C. C. & Ellenberger, T. (2001). *Proc. Natl. Acad. Sci. USA*, **98**, 9557–9562.
Kerr, I. D., Wadsworth, R. I. M., Blankenfeldt, W., Staines, A. G., White, M. F. & Naismith, J. H. (2001). *Acta Cryst.* D**50**, 760–763.
Lohman, T. M. & Ferrari, M. E. (1994). *Annu. Rev. Biochem.* **63**, 527–570.
Matsumoto, T., Morimoto, Y., Shibata, N., Kinebuchi, T., Shimamoto, N., Tsukihara, T. & Ysuoka, N. (2000). *J. Biochem.* **127**, 329–335.
Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.
Ng, J. D. & McPherson, A. (1989). *J. Biomol. Struct. Dynam.* **6**, 1071–1076.
Ollis, D., Brick, P., Abdel-Maguid, S. S., Murthy. K., Chase, J. W. & Steitz, T. A. (1983). *J. Mol. Biol.* **170**, 797–801.
Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–325.
Purnapatre, K. & Varshney, U. (1999). *Eur. J. Biochem.* **264**, 591–598.
Raghunathan, S., Kozlov, A. G., Lohman, T. M. & Waksman, G. (2000). *Nature Struct. Biol.* **145**, 75–104.
Raghunathan, S., Ricard, C. S., Lohman, T. M. & Waksman, G. (1997). *Proc. Natl Acad. Sci. USA*, **94**, 6652–6657.
Reddy, M. S., Guhan, N. & Muniyappa, K. (2001). In the press.
Shamoo, Y., Friedman, A. M., Parsons, M. R., Konigsberg, W. H. & Steitz, T. A. (1995). *Nature (London)*, **376**, 362–366.
Suck, D. (1997). *Nature Struct. Biol.* **4**, 161–165.
Vries, J. de, Genshel, J., Urbanke, C., Thole, H. & Wackernagel, W. (1994). *Eur. J. Biochem.* **224**, 613–622.
Webster, G., Genschel, J., Curth, U., Urbanke, C., Kang, C. & Hilgenfeld, R. (1997). *FEBS Lett.* **411**, 313–316.
Yang, C., Curth, U., Urbanke, C. & Kang, C. (1997). *Nature Struct. Biol.* **4**, 153–157.