Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

X-ray and molecular-dynamics studies on *Mycobacterium leprae* single-stranded DNA-binding protein and comparison with other eubacterial SSB structures

Prem Singh Kaushal,[a] Pawan Singh,[b] Alok Sharma,[a] K. Muniyappa[b] and M. Vijayan[a]*

[a]Molecular Biophysics Unit, Indian Institute of Science, Bangalore 560 012, India, and [b]Department of Biochemistry, Indian Institute of Science, Bangalore 560 012, India

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

Received 23 June 2010
Accepted 11 August 2010

PDB References:
Mycobacterium leprae SSB, form I, 3afp; form II, 3afq.

The crystal structures of two forms of *Mycobacterium leprae* single-stranded DNA-binding protein (SSB) have been determined at 2.05 and 2.8 Å resolution. Comparison of these structures with the structures of other eubacterial SSBs indicates considerable variation in their quaternary association, although the DNA-binding domains in all of them exhibit the same OB-fold. This variation has no linear correlation with sequence variation, but could be related to variation in protein stability. Molecular-dynamics simulations have been carried out on tetrameric molecules derived from the two forms and the prototype *Escherichia coli* SSB and the individual subunits of both proteins. Together, the X-ray studies and molecular-dynamics simulations yield information on the relatively rigid and flexible regions of the molecule and on the effect of oligomerization on flexibility. The simulations provide insight into the changes in subunit structure on oligomerization. They also provide insight into the stability and time evolution of the hydrogen bonds/water bridges that connect the two pairs of monomers in the tetramer.

1. Introduction

A number of studies have shown that DNA transactions such as replication, recombination and repair involve the separation of duplex DNA into its constituent single-stranded DNA (ssDNA). Single-stranded DNA-binding proteins (SSBs) are involved in the protection of ssDNA from the formation of aberrant secondary structures, genotoxic compounds and nucleolytic degradation. The crystal structures of human mitochondrial SSB (PDB code 3uul; Yang *et al.*, 1997) and of *Escherichia coli* SSB (EcSSB; PDB code 1kaw; Raghunathan *et al.*, 1997) were published simultaneously. The two proteins are homotetrameric and have similar structures. Each subunit of the tetramer contains a well ordered N-terminal DNA-binding domain and a disordered C-terminal tail which is involved in interactions with other proteins (Shereda *et al.*, 2008). The C-terminal tail is disordered in the crystal structures of all SSB proteins reported to date. The crystal structures of *Mycobacterium tuberculosis* SSB (MtSSB; PDB code 1ue1) and *M. smegmatis* SSB (MsSSB; PDB code 1x3e) were then reported from this laboratory (Saikrishnan *et al.*, 2003, 2005) as part of a larger program of structural studies of mycobacterial proteins (Vijayan, 2005; Krishna *et al.*, 2006; Selvaraj *et al.*, 2007; Kaushal *et al.*, 2008; Chetnani *et al.*, 2010). The subunit structures of these proteins are similar to that of

the prototype EcSSB, but there are significant differences in their quaternary association. In the meantime, the crystal structure of *Deinococcus radiodurans* SSB (DrSSB) became available (PDB code 1se8; Bernstein *et al.*, 2004). DrSSB is a dimeric protein, but each subunit contains two OB domains, so that the overall geometry of the molecule is similar to the known tetrameric SSB structures. Since then, crystal structures of SSB from *Thermus aquaticus* (TaSSB; PDB code 2fxq; Jędrzejczak *et al.*, 2006), *T. thermophilus* (TtSSB; PBD code 2cwa; S. Satoh, S. Yokoyama & S. Kuramitsu, unpublished work), *Thermotoga maritima* (TmSSB; PDB code 1z9f; DiDonato *et al.*, 2006), *Mycoplasma pneumoniae* (MpSSB; PDB code 2hql; Das *et al.*, 2007) and *Streptomyces coelicolor* (ScSSB; PDB code 3eiv; Štefanić *et al.*, 2009) have been determined. The crystal structures of DNA-bound SSBs from *E. coli* (PDB code 1eyg; Raghunathan *et al.*, 2000) and *Helicobacter pylori* (HpSSB; PDB code 2vw9; Chan *et al.*, 2009) are also available. These structures exhibit varying degrees of similarity to those of the SSBs mentioned earlier.

Although MtSSB and MsSSB have similar structures, as expected, subtle variations in their quaternary association were clearly discernible. One of the variations involves the presence of either direct hydrogen bonds or water bridges between the two subunits. To further explore this variability and to strengthen the structural information on mycobacterial SSB proteins, we determined the three-dimensional structure of *Mycobacterium leprae* SSB (MlSSB). One of the crystal forms has water bridges between the two subunits, while the other presents an intermediate situation that involves both direct hydrogen bonds and water bridges. A comparative study involving these structures and those of other eubacterial SSB proteins led to valuable insights, especially into the relationship between quaternary association and protein stability. The X-ray results were supplemented by molecular-dynamics (MD) simulations on tetrameric molecules and on individual subunits derived from the crystal structures of MlSSB and EcSSB. The simulations were particularly useful in elucidating the changes in the subunit structure brought about by oligomerization. They also provided information on the dynamics of the hydrogen-bonded/water-mediated inter-subunit interactions referred to earlier. Together, the X-ray results, including those reported here, and the simulations lead to a reasonably comprehensive picture of the structure and dynamics of SSB.

2. Materials and methods

2.1. Cloning, expression and purification

The open reading frame corresponding to the *M. leprae ssb* gene was PCR-amplified with gene-specific primers (forward primer 5′-AGGGAGGAA**CATATG**GCTGGCGACACC-3′; reverse primer 5′-CTTTCCGTT**CTCGAG**CTAGAAGGGG-GGTTC-3′) using cosmid MLCY454 as the DNA template. The PCR product was cloned into a pET21a expression vector at the *Nde*I/*Xho*I restriction site downstream of the phage T7 promoter. The resulting recombinant plasmid was designated pMLSSB and the identity of this plasmid was confirmed by restriction analysis and DNA sequencing. The protein was overexpressed by transforming the pMLSSB recombinant plasmid into BL21 (DE3) pLysS host strain. *M. leprae* SSB was purified as described previously (Reddy *et al.*, 2001). The protein was purified to >98% homogeneity and its concentration was determined by the dye-binding method using bovine serum albumin as a standard (Bradford, 1976).

Table 1
Crystal data and data-collection, refinement and model statistics.

Values in parentheses are for the highest resolution shell.

	Form I	Form II
Space group	$P3_121$	$P3_221$
Unit-cell parameters (Å)		
a	78.96	102.48
b	78.96	102.48
c	80.26	120.76
Unit-cell volume (Å^3)	433068.0	1098422.6
V_M (Å^3 Da^{-1})	2.04	2.58
Solvent content (%)	39.6	52.4
No. of subunits in the asymmetric unit	2	4
Resolution range (Å)	28.14–2.05 (2.1–2.05)	30.0–2.8 (2.87–2.8)
No. of observed reflections	72178 (1116)	44491 (1130)
No. of unique reflections	18540 (1824)	18449 (1810)
Completeness (%)	99.8 (99.6)	99.7 (97.8)
Multiplicity	3.9	2.4
Average $I/\sigma(I)$	21.6 (3.1)	15.3 (2.0)
R_{merge}† (%)	6.0 (46.1)	5.9 (51.7)
Refinement and model statistics		
R factor (%)	20.4	21.1
R_{free}‡ (%)	23.2	23.5
R.m.s. deviation from ideal		
Bond lengths (Å)	0.018	0.018
Bond angles (°)	1.5	1.5
Ramachandran plot, residues in (%)		
Core regions	89.5	83.7
Allowed regions	6.8	14.7
Generously allowed regions	3.7	1.6
Disallowed regions	0.0	0.0
No. of protein atoms	1687	3165
No. of water molecules	217	154
No. of cadmium ions	1	—
No. of glycerol molecules	1	—

† $R_{merge} = \sum_{hkl}\sum_i |I_i(hkl) - \langle I(hkl)\rangle| / \sum_{hkl}\sum_i I_i(hkl)$, where $I_i(hkl)$ is the *i*th observation of reflection *hkl* and $\langle I(hkl)\rangle$ is the weighted average intensity for all observations *i* of reflection *hkl*. ‡ R_{free} was calculated from a randomly selected 5% of unique reflections that were omitted from structure refinement.

2.2. Crystallization

MlSSB was screened for crystallization using Crystal Screen, Crystal Screen 2 and Index kits from Hampton Research. Crystals appeared in two forms. Approximately cube-shaped form I crystals of dimensions 0.2 × 0.2 × 0.25 mm appeared after 7 d in a hanging drop consisting of 4 µl 7 mg ml^{-1} protein in 20 m*M* Tris–HCl pH 7.4 and 0.5 *M* NaCl and 1 µl reservoir solution made up of 1 *M* sodium acetate, 500 m*M* NaCl, 30 m*M* $CdSO_4$ and 20 m*M* Tris–HCl pH 7.5. Crystals were transferred to reservoir solution containing 30% glycerol as a cryoprotectant before flash-freezing. Form II crystals were prepared using the microbatch method from a solution containing 3 µl 7 mg ml^{-1} protein in the same buffer and 3 µl precipitant solution made up of 0.1 *M* HEPES

pH 7.5 and 3 *M* NaCl. The crystals were hexagonal prisms of approximate dimensions 0.2 × 0.2 × 0.4 mm.

2.3. X-ray data collection and processing

Diffraction data were collected at low temperature (100 K) using a MAR Research image-plate system (345 mm diameter) with Osmic mirrors and a Rigaku RU-200 rotating-anode X-ray generator. Intensity data were processed and scaled using *DENZO* and *SCALEPACK* from the *HKL* program package (Otwinowski & Minor, 1997). Crystal form I belonged to space group $P3_121$, with unit-cell parameters $a = b = 78.96$, $c = 80.26$ Å, and diffracted to a resolution of 2.05 Å. Crystal form II belonged to space group $P3_221$, with unit-cell parameters $a = b = 102.48$, $c = 120.76$ Å, and diffracted to a resolution of 2.8 Å. Intensities were converted to structure-factor amplitudes using the program *TRUNCATE* from *CCP*4 (Collaborative Computational Project, Number 4, 1994). Both data sets were tested for twinning using the program *DETWIN* from *CCP*4. The data from crystal form II indicated twinning (twin fraction of 0.356 for twin operator $-h, -k, l$), while those from crystal form I did not.

2.4. Structure solution and refinement

The form I crystals were isomorphous to the MsSSB crystals (PDB code 1x3e). The MsSSB structure was taken as the initial model for refinement of form I using the program *CNS* v.1.1 (Brünger *et al.*, 1998). MlSSB and MsSSB have sequence identities of 87% for the full-length protein and 95% for the DNA-binding domain. Rigid-body and positional refinement, simulated-annealing and individual *B*-factor refinement were successively carried out with model building using *Coot* (Emsley & Cowtan, 2004) interspersed between refinement cycles. Cadmium ions were defined on the basis of very strong electron density in both $2F_o - F_c$ and $F_o - F_c$ maps. Peaks in these maps were also used to identify water molecules. The final cycles of refinement were carried out using the program *REFMAC* (Murshudov *et al.*, 1997) in *CCP*4. In this step, TLS refinement (Winn *et al.*, 2003) was carried out. Each subunit was divided into four TLS groups. Group I represents the molecular core, consisting of residues 2–21, 28–34, 53–85 and 98–120, while groups II, III and IV represent loops consisting of residues 22–27, 5–52 and 86–97, respectively. The refined subunit of form I was used as the search model in the molecular-replacement solution of form II. A Matthews coefficient (Matthews, 1968) of 2.04 Å^3 Da^{-1} corresponded to the presence of four

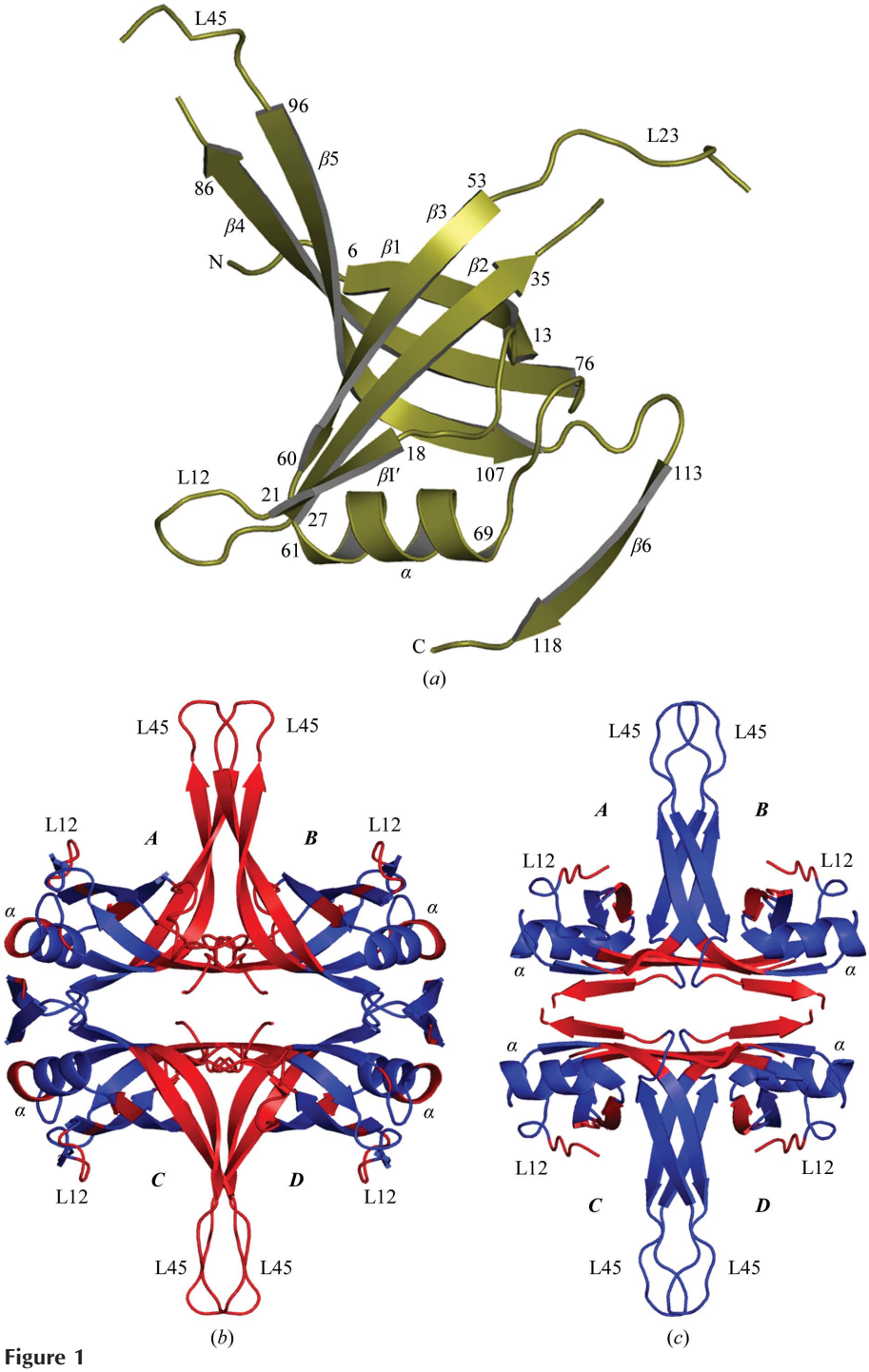

Figure 1
(*a*) Structure of the DNA-binding domain in subunit *A* of form I MlSSB. The secondary-structural elements are numbered as in EcSSB (Raghunathan *et al.*, 1997). Tetrameric molecules of (*b*) MlSSB and (*c*) EcSSB are shown. The relatively rigid and flexible regions are shown in blue and red, respectively. See text for details.

subunits in the asymmetric unit with 40.0% solvent content. *Phaser* (Storoni *et al.*, 2004) gave the best solution in space group $P3_221$, with a log-likelihood gain (LLG) of 1298 and a Z score of 25 for four crystallographically independent subunits. A twinning correction was applied throughout refinement. The structure was refined in a manner similar to that used for form I. During the initial refinement cycles, NCS restraints were used with a force constant of 1256 kJ mol^{-1}, which was finally relaxed to 837 kJ mol^{-1}. NCS restraints were not applied to the terminal and loop regions. Both forms refined to acceptable values of R and $R_{\rm free}$ and with good stereochemistry. The N-terminal residues and a few residues in the L23 and L45 loops are not defined in all of the subunits. Data-collection, refinement and model statistics are given in Table 1.

2.5. Molecular-dynamics simulation

The software package *GROMACS* v.3.3.1 (van der Spoel *et al.*, 2005) was used to perform molecular-dynamics (MD) simulations employing the OPLS-AA/L force field (Jorgensen & Maxwell, 1996). Protein molecules as derived from the relevant crystal structures were used as starting models. An octahedral simulation box was generated around the protein molecule using the *editconf* module of the *GROMACS* package. The dimensions were selected so that the minimum distance between the protein molecule and the edge of the box was 7.5 Å. The box volumes were roughly 72×10^4 and 34×10^4 Å^3 in the case of tetramers and monomers, respectively. Protein models were solvated with the TIP4P water model and wherever necessary the system was neutralized with Cl^- ions using the program *genbox*. Energy calculation was performed in cycles of one step of steepest-descent minimization followed by 1000 steps of conjugate-gradient minimization. Calculations were terminated when the change in energy in the last cycle was less than 1 kJ mol^{-1} nm^{-1}. The energy-minimized system was further subjected to solvent equilibration by position-restrained dynamics of 10 ps. Solvent molecules were allowed to move, whereas protein atoms were restrained. Simulations were performed under NPT conditions using Parrinello–Rahman isotopic pressure coupling (τ_p = 0.5 ps) to 100 kPa and Nose–Hoover temperature coupling (τ_p = 0.5 ps) to 300 K. The Particle Mesh Ewald (PME) method (Darden *et al.*, 1993) with a grid spacing of 1.5 Å was used to evaluate long-range electrostatic interactions with a cutoff of 12 Å. Long-range van der Waals interactions were calculated employing a cutoff of 15 Å. Periodic boundary conditions (PBC) were used throughout the simulation. Bonds were constrained with the *LINCS* algorithm (Hess *et al.*, 1997). Integrations were performed at a 2 fs time step. A dielectric constant of unity was used. MD simulations were performed for a time period of 20 ns.

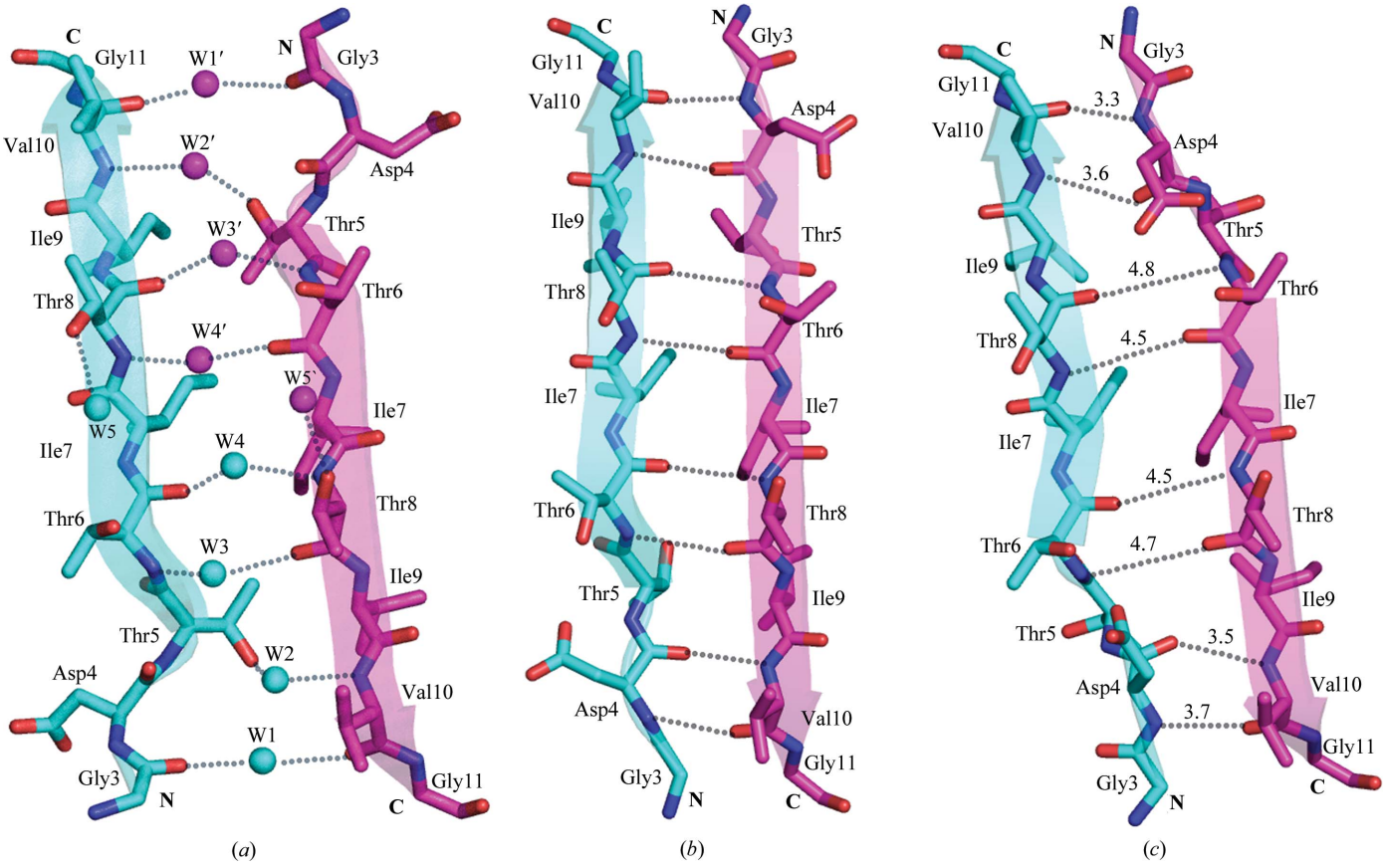

Figure 2
Interactions between monomers (*a*) in form I MlSSB, (*b*) in form II MtSSB and (*c*) at the *AB* interface in form II MlSSB. The distances in (*c*) indicate that the four inner interactions are water bridges.

2.6. Analysis of structures

The refined model was evaluated using *PROCHECK* (Laskowski *et al.*, 1993). Sequence alignment was performed using *ClustalW*2 (Larkin *et al.*, 2007). Structures were superposed using *ALIGN* (Cohen, 1997). MD simulation data were analyzed with various tools available in the *GROMACS* suite and scripts and programs developed in the laboratory. Graphs were prepared using *Xmgr* (Paul J. Turner, Centre for Coastal and Land-Margin Research, Oregon Graduate Institute of Science and Technology, Beaverton, Oregon). *HBPLUS* was used to identify hydrogen bonds (McDonald & Thornton, 1994). Criteria of a donor–acceptor distance less than or equal to 3.2 Å and a donor–hydrogen–acceptor angle greater than or equal to 120° were used to delineate hydrogen bonds. Limits of 3.6 Å and 90° were used for the X-ray structures to allow for their limited resolutions. The buried surface area was taken as the difference between the sum of the accessible surface areas of the components and that of the complex. Accessible surface areas were calculated employing *NACCESS* (Hubbard & Thornton, 1993) using a probe size of 1.4 Å. The relatively rigid and flexible regions were estimated using the *Error-inclusive Structure Comparison and Evaluation Tool* (*ESCET*; Schneider, 2002). Figures were generated using the program *PyMOL* (DeLano, 2002).

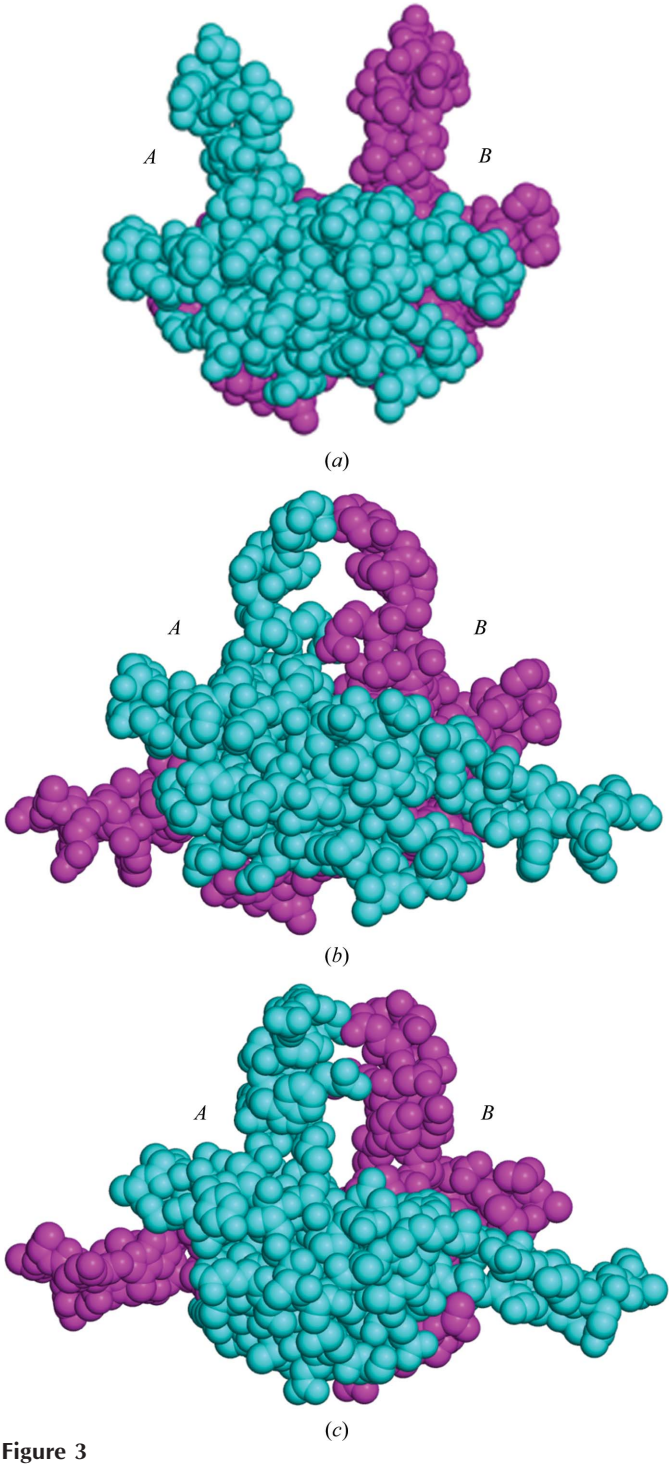

Figure 3
The mutual orientation of the monomers in the dimer in (*a*) ScSSB, (*b*) MlSSB and (*c*) EcSSB. The two monomers are shown in different colours.

3. Results and discussion

3.1. Overall structure of *M. leprae* SSB and comparison with other mycobacterial SSB proteins

Like most other eubacterial SSB proteins, MlSSB is tetrameric in both of the crystal forms reported here, with the DNA-binding domain in each subunit exhibiting an OB-fold. The C-terminal stretch beyond residue 120 is disordered in both cases. The globular core of the DNA-binding domain is made up of a six-stranded β-barrel capped by an α-helix. Three loops, L12, L23 and L45, extend out of the globular core. Like other mycobacterial SSB proteins, MlSSB has an additional β-strand, designated β6 in Fig. 1(*a*), which is involved in stabilizing the tetrameric structure through strand-swapping. A whole tetramer with noncrystallographic 222 symmetry exists in form II of MlSSB. One of the twofold axes is crystallographic in form I such that a dimer (subunits *A* and *C* in Fig. 1*b*) exists in the asymmetric unit in the crystal form.

MtSSB, MsSSB and MlSSB are homologous, with a high sequence identity (87–92%). The identity is higher still when the DNA-binding domain alone is considered. Consequently, the proteins from the three mycobacterial sources have nearly identical tertiary and quaternary structures. The loops exhibit some variability and differing extents of disorder. A major characteristic feature of mycobacterial SSB proteins is the additional stabilization of the tetramer through the involvement of β6, which clamps the two dimers together at the ends. In spite of the similarity among them, a subtle variability exists among the structures in relation to the inter-protomer interface within the dimer. In form I of MlSSB, form I of MtSSB and MsSSB the two subunits are interconnected by water bridges (Fig. 2*a*). Direct hydrogen bonds exist between them in form II of MtSSB (Fig. 2*b*), while an intermediate situation exists in form II of MlSSB (Fig. 2*c*). In this structure, of the eight hydrogen-bonded interactions between the two subunits, the four in the middle appear to involve water bridges, while the outer four involve direct hydrogen bonds in the *AB* dimer (Fig. 2*c*). All eight interactions are direct hydrogen bonds in the *CD* dimer.

3.2. Comparison with known eubacterial SSB protein structures

Among the known eubacterial SSB protein structures, that from *S. coelicolor* (ScSSB) is closest to the mycobacterial SSB proteins in terms of sequence, with a sequence identity of 64–70%. This similarity in sequence is also reflected in the r.m.s. deviations in C^{α} positions on superposition of the relevant structures. Strand β6 and the clamps involving this strand also exist in ScSSB. All eight water bridges exist between the two subunits in the dimers in the structure of ScSSB. The OB-fold SSB protein of known structure that is most distant from the mycobacterial SSB proteins is that of the dimeric SSB from *Mycoplasma pneumoniae* (MpSSB), with a sequence identity as low as 2–7%. Despite the low sequence identity, the overall structure of the domain remains the same as in the mycobacterial and other similar SSB proteins. All eight interactions between two subunits also involve direct hydrogen bonds. The most striking difference between MpSSB and the other eubacterial SSB proteins considered here concerns oligomerization. MpSSB is a dimer, with each subunit involving an OB-fold. The other SSB proteins contain four OB-folds distributed among four or two subunits. However, the structure of the MpSSB dimer is very similar to one half of the SSB molecules from other sources.

Of the other known SSB protein structures, EcSSB and HpSSB are bacterial, HsSSB is eukaryotic but mitochondrial and TaSSB and TtSSB are from thermophilic bacteria. *T. maritima* is a bacterium, but its genome exhibits some archaeal character. *D. radiodurans* is an unusual extremophilic bacterium. The correlation between the similarity of the species and that of the SSB sequence is only approximate. The same is true of the correlation involving r.m.s. deviations in C^{α} positions when the corresponding DNA-binding domains are superimposed. All of the bacterial SSB proteins, except for DrSSB, TaSSB, TtSSB and MpSSB, are tetrameric, with each subunit involving one OB-fold. DrSSB and the two *Thermus* SSBs TaSSB and TtSSB are dimeric, but each subunit corresponding to the *AB* dimer in the tetrameric SSBs contains two OB-folds. Thus, the geometries of the dimeric SSBs are very similar to those of the tetrameric SSBs. Interestingly, the sequence identity between the two domains in each subunit, which ranges between 31 and 34%, in dimeric SSB proteins is not greatly different from that between these domains and SSB proteins from other species. In spite of these seemingly irrational sequence variations among SSB proteins from different species, the structure of the DNA-binding domains in four-domain SSB proteins remains nearly the same, with a maximum r.m.s. deviation of 0.9 Å between domains.

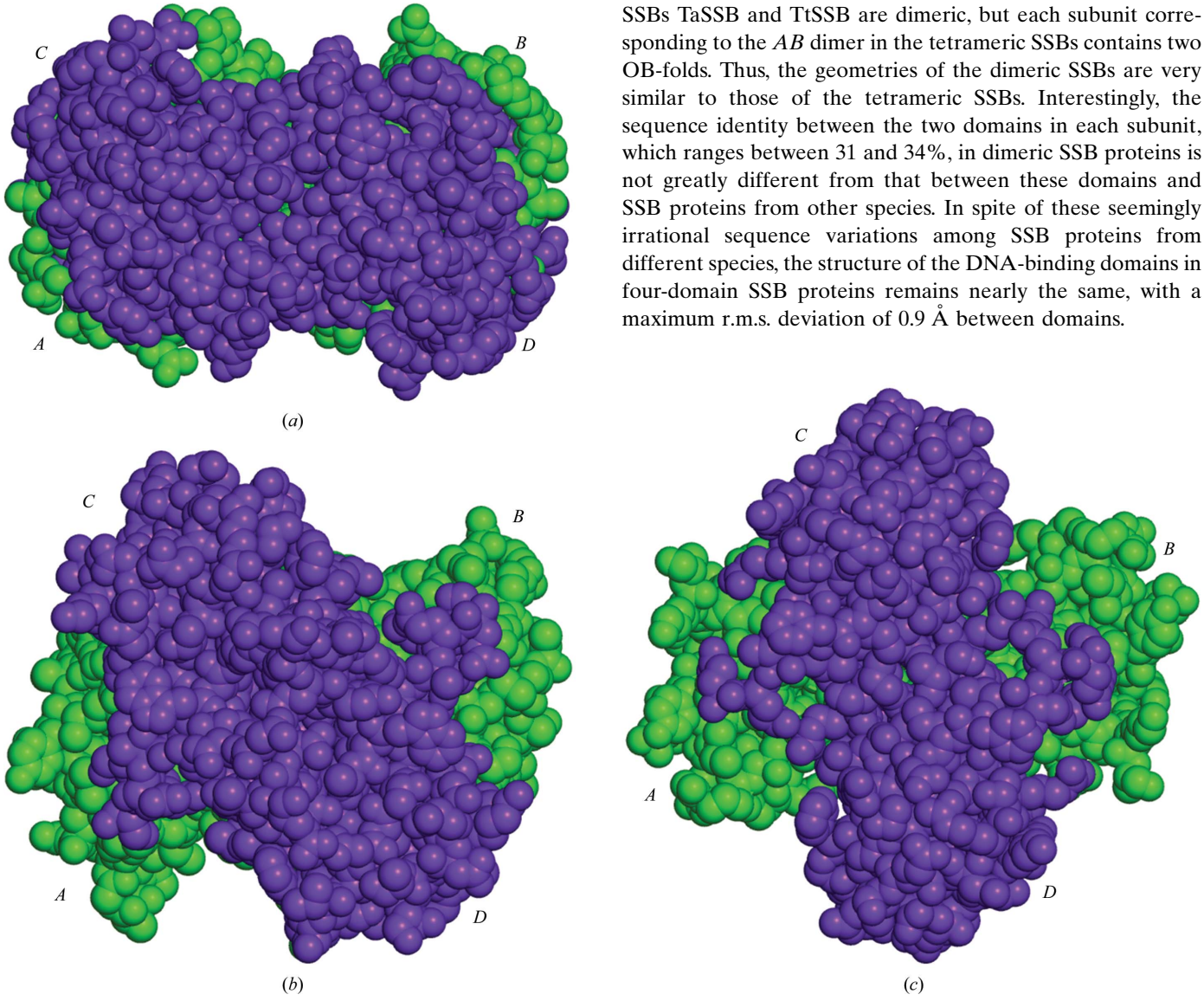

Figure 4
The mutual orientation of the dimers in the tetramer in (*a*) MlSSB, (*b*) EcSSB and (*c*) HpSSB. The two dimers are shown in different colours.

Table 2
Surface area (Å^2) buried on oligomerization.

The nonpolar component is given in parentheses.

Source	PDB code	Tetramer	*AB*	*AC*	*AD*
Mycobacterium leprae form I	3afp	9048 (5289)	1805 (979)	2805 (1724)	28 (17)
M. leprae form II	3afq	9630 (5935)	2074 (1226)	2664 (1689)	92 (34)
M. tuberculosis form I	1ue1	8526 (5167)	1641 (1002)	2428 (1520)	220 (55)
M. tuberculosis form II	1ue6	8568 (5150)	1690 (950)	2450 (1561)	152 (48)
M. smegmatis form I	1x3e	8932 (5446)	1673 (896)	2770 (1734)	34 (20)
Streptomyces coelicolor	3eiv	9312 (5661)	1654 (916)	2837 (1776)	16 (12)
Escherichia coli	1kaw	5935 (3335)	2319 (1135)	410 (335)	637 (424)
E. coli + DNA	1eyg	6490 (3536)	2333 (1180)	417 (355)	659 (378)
Homo sapiens	3ull	7853 (4240)	2770 (1482)	588 (392)	830 (407)
Helicobacter pylori + DNA	2vw9	5681 (3766)	2362 (1598)	500 (360)	105 (94)
Thermotoga maritima	1z9f	7459 (4770)	2594 (1599)	387 (369)	1081 (654)

3.3. Variability in quaternary structure

The *AB* dimer in tetrameric SSB proteins and the subunits containing two DNA-binding domains in DrSSB, TaSSB and TtSSB have essentially the same structure in all of the crystals. However, some rotation of one subunit with respect to the other is observed when comparing *AB* dimers from different species. As illustrated in Fig. 3, ScSSB and EcSSB represent the two extreme situations among the microbial SSBs (Figs. 3*a* and 3*c*), while mycobacterial SSB proteins lie between the two (Fig. 3*b*). The variability in the mutual orientation of the two dimers in the tetramer is still more striking (Fig. 4). The mycobacterial proteins and ScSSB represent one extreme with a maximum overlap between the interfaces of the two dimers (Fig. 4*a*). The clamping together of the two dimers using β6 apparently locks the molecule in this arrangement. The other extreme is represented by HpSSB (Fig. 4*c*). The arrangement in HsSSB and EcSSB is in between the two but is closer to the latter (Fig. 4*b*).

Mycobacterial SSB proteins and ScSSB, and in particular their DNA-binding domains, form a highly homologous group. They also have substantially the same quaternary structure, which is locked in position in each case by clamps involving β6. The variation of the quaternary arrangements in other SSB proteins, both among themselves and with respect to those in mycobacterial SSB proteins, could not be rationalized in terms of variation in sequence alone. As can be seen from the data presented in Table 2, the surface area buried on tetramerization is largest in the mycobacterial SSB proteins and ScSSB. The other bacterial SSB proteins and HsSSB exhibit a lower burial of surface area on tetramerization. The situation in DrSSB and the two *Thermus* SSB proteins TaSSB and TtSSB is not comparable to that in other SSB proteins as two DNA-binding domains are covalently linked in them.

The contribution of different components to the total buried surface area also varies substantially. The surface area buried on dimerization is uniformly lower in mycobacterial SSB proteins and ScSSB than in other SSB proteins. This can be seen as a consequence of the difference in the mutual orientation of subunits *A* and *B*. The situation is reversed when dealing with the major dimer–dimer interface (*AC* and the equivalent *BD*). This again is clearly a consequence of the difference in the mutual orientation of the two dimers in the tetramer. The surface area buried across subunits *A* and *D* (and *B* and *C*) is comparatively small in all cases. Interactions across this interface involve C-terminal residues and the actual value of the area buried depends upon the orientation and the degree of order of these residues. When all the components are added, the total surface area buried on tetramerization, for both polar and nonpolar individually and together, is distinctly higher in mycobacterial SSB proteins and ScSSB, which are the only SSB proteins in which the two dimers are clamped together through strand-swapping. This observation is consistent with the demonstrated higher stability of MtSSB in comparison with EcSSB in the presence of guanidium hydrochloride (Handa *et al.*, 2000).

3.4. MD simulations: general features

MD simulations were carried out on three tetrameric models of SSB proteins: two derived from the two crystal forms of MlSSB reported here and one derived from the crystal structure of EcSSB (PDB code 1kaw). MlSSB and EcSSB represent two modes of quaternary association. Furthermore, mycobacterial SSB proteins, which share the same quaternary structure with EcSSB, are among the best-studied SSB proteins. Although all eubacterial SSB proteins have nearly the same dimeric structure, there are subtle differences between them. Many have direct β-sheet-type hydrogen bonds between the two monomers, as is the case in EcSSB. These hydrogen bonds are replaced by water bridges in some, including form I of MlSSB. The molecule in form II of MlSSB represents an intermediate situation. Thus, the tetramers chosen for the simulation also represent three slightly different modes in terms of inter-monomer interactions in the dimer. In order to explore structural changes that accompany oligomerization, simulations were also carried out on the subunits of MlSSB and EcSSB. The time-evolution of the root-mean-square deviations indicated that the system stabilized within 10 ns (Supplementary Fig. S1[1]). All of the parameters in the following discussion are derived from the results of simulations in the 10–20 ns range.

The difference between the tetramers derived from the two forms of MlSSB corresponds to the water bridges between the two subunits in the *AB* dimer. Therefore, except in relation to this feature, the simulation involving the tetramer derived from crystal form I is used in the discussion here. The root-mean-square fluctuation (r.m.s.f.) in C^{α} positions in this simulation, together with those in the simulation of a subunit of MlSSB, is shown in Fig. 5(*a*). Also shown in the figure are the average *B* values of the same atoms in the crystal structure.

[1] Supplementary material has been deposited in the IUCr electronic archive (Reference: BE5153). Services for accessing this material are described at the back of the journal.

The six subunits in the two crystal structures were used to estimate the relative rigid and flexible regions of the subunit using the *Error-inclusive Structure Comparison and Evaluation Tool* (*ESCET*; Schneider, 2002). The parameter σ was chosen so that the residues are nearly equally divided between the two regions. The rigid and flexible regions thus delineated are indicated in Fig. 5(*a*). Also indicated in the figure are the secondary-structural features and the regions involved in the *AB* and *AC* interfaces. The corresponding data for EcSSB are illustrated in Fig. 5(*b*). The relatively rigid and flexible regions in this case were delineated from the four subunits in the tetrameric molecule in the crystal structure. Interestingly, in MlSSB the monomer–monomer (*AB* and *CD*) interface is relatively flexible (Fig. 1*b*). Presumably, this flexibility is responsible for the variability in the interactions in terms of hydrogen bonds and water bridges at the interface. In contrast, the dimer–dimer interface is relatively flexible in EcSSB (Fig. 1*c*). In EcSSB the two dimers are not clamped together, unlike in mycobacterial SSB proteins.

The reasonable correlation among the r.m.s.f.s obtained from simulations, the *B* values obtained from X-ray studies and the estimated relatively rigid and flexible regions of the molecule on the basis of comparison among different subunits in crystal structures (Fig. 5) is a source of reassurance. The correlation breaks down in the regions around the L45 loop in

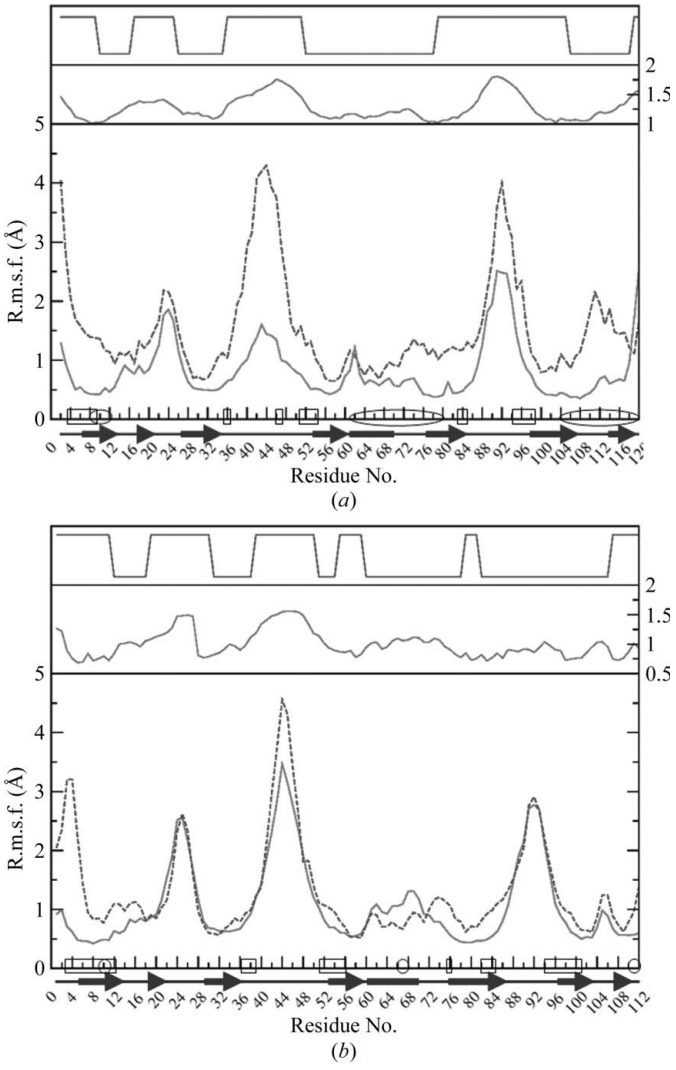

Figure 5
(*a*) Flexibility in MlSSB along the polypeptide chain. The lower panel illustrates the r.m.s.f. during simulations on an individual subunit (upper curve) and the tetrameric molecule (lower curve). Average values of the displacement derived from crystallographic *B* values are shown in the middle panel. The rigid (low level) and the flexible (high level) stretches are indicated in the top panel. The rectangles and ellipses on the abscissa indicate residues in the monomer–monomer and dimer–dimer interfaces, respectively. The secondary-structure elements are indicated just above the residue numbers at the bottom. (*b*) Flexibility in EcSSB along the polypeptide chain. Descriptions of the contents are as in Fig. 5(*a*).

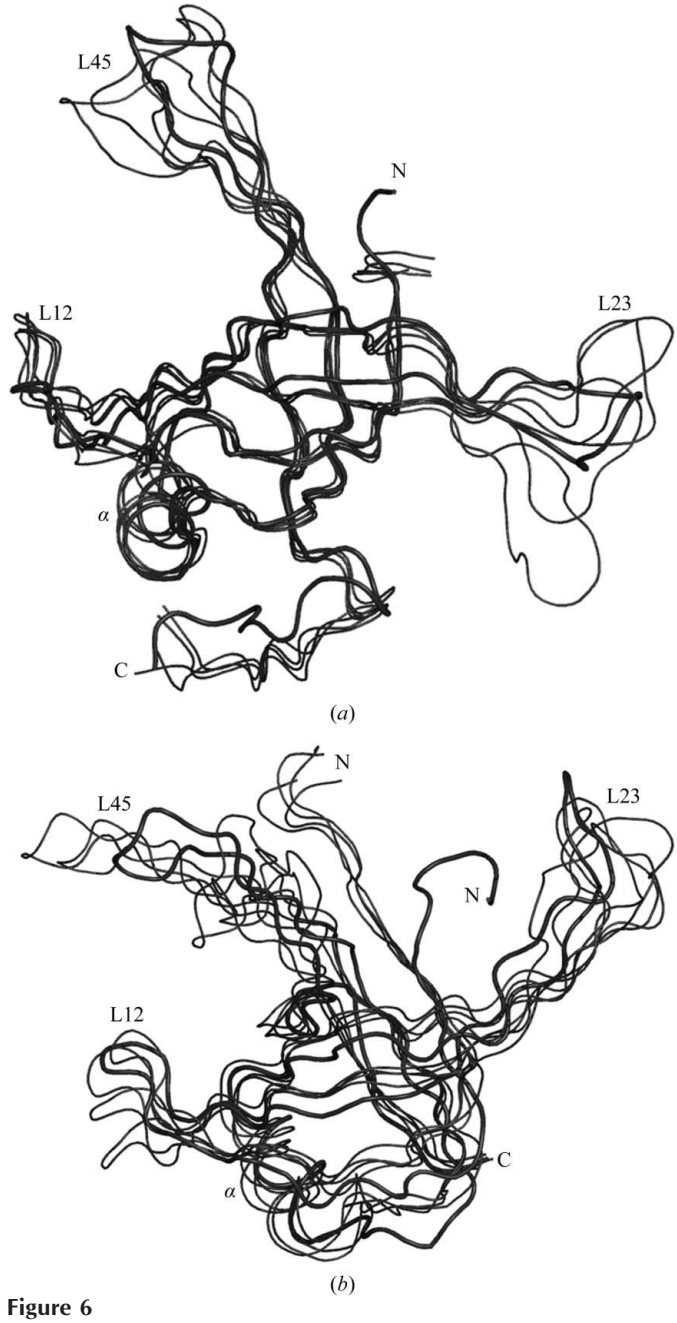

Figure 6
Structures corresponding to the peak in the population distribution in the subunit simulation (thick line) and the four subunits in the tetrameric simulation (thin lines) in (*a*) MlSSB and (*b*) EcSSB.

EcSSB. Simulation indicated this region to be relatively flexible, whereas X-ray results indicate it to be relatively rigid. This is because this loop is involved in intermolecular interactions in the crystal structure and hence is locked in the crystal. Understandably, r.m.s.f. values are in general lower in the tetramer than in the individual subunit. The reduction in the r.m.s.f. on oligomerization is more pronounced in MlSSB than in EcSSB. This correlates well with the higher stability of mycobacterial SSB proteins on account of strand-swapping and the larger surface area buried on tetramerization. Furthermore, the reduction is particularly pronounced in the regions of the molecule that are involved in oligomerization in the case of MlSSB. In general, again understandably, loops are more flexible than strands.

3.5. Change in subunit structure on oligomerization

MlSSB and EcSSB exist as tetramers in their crystal structures, which do not provide information on the changes brought about in the structure of the subunit as a consequence of oligomerization. However, independent MD simulations of the tetrameric molecule and the subunit can be used for this purpose. The regions of the subunit that are involved in the formation of the *AB* (and *CD*) dimer are essentially the same in MlSSB and EcSSB, although the number of residues that take part in the interaction is larger and the association is therefore tighter in EcSSB. On the whole, in addition to the N-terminal and C-terminal stretches, the regions that are consistently involved in oligomerization are those containing L23 and L45. It turns out that these regions exhibit substantial changes on oligomerization (Fig. 6). They also show greater flexibility and variability in their structures (Figs. 5 and 6). The region involving loop L12, which is not involved in oligomerization, shows less flexibility.

The N-terminal stretch is extensively involved in oligomerization in both proteins. However, this stretch behaves somewhat differently in the simulation of independent subunits. The most conspicuous role of β1 in this stretch is in the hydrogen bonds or water bridges that are formed during dimerization. They also interact across the other two interfaces of the tetramer in both proteins. In the independent subunits, however, this stretch moves away from the body of the subunit in EcSSB, while it closely clings to the body in MlSSB (Fig. 7). This appears to be on account of the additional hydrogen bonds, as illustrated in Fig. 7(*a*), that are generated on dissociation of the tetramer. One of them involves the side-chain amino group of Gln85. The residue at this location in EcSSB is threonine, which obviously cannot form the hydrogen bond. Presumably, the remaining additional hydrogen bonds are not sufficient to hold the stretch in position in EcSSB. The N-terminal stretch, which is not otherwise constrained in the independent subunit, thus moves away from the body of the subunit. As indicated earlier, the C-terminal stretch locks the two dimers together in mycobacterial SSB proteins. The unlocking of the stretch on the dissociation of the tetramer in the free subunit, however, does not cause its free movement. This is because of the formation of new hydrogen bonds, as illustrated in Fig. 7(*b*), with the body of the subunit on the dissociation of the tetramer.

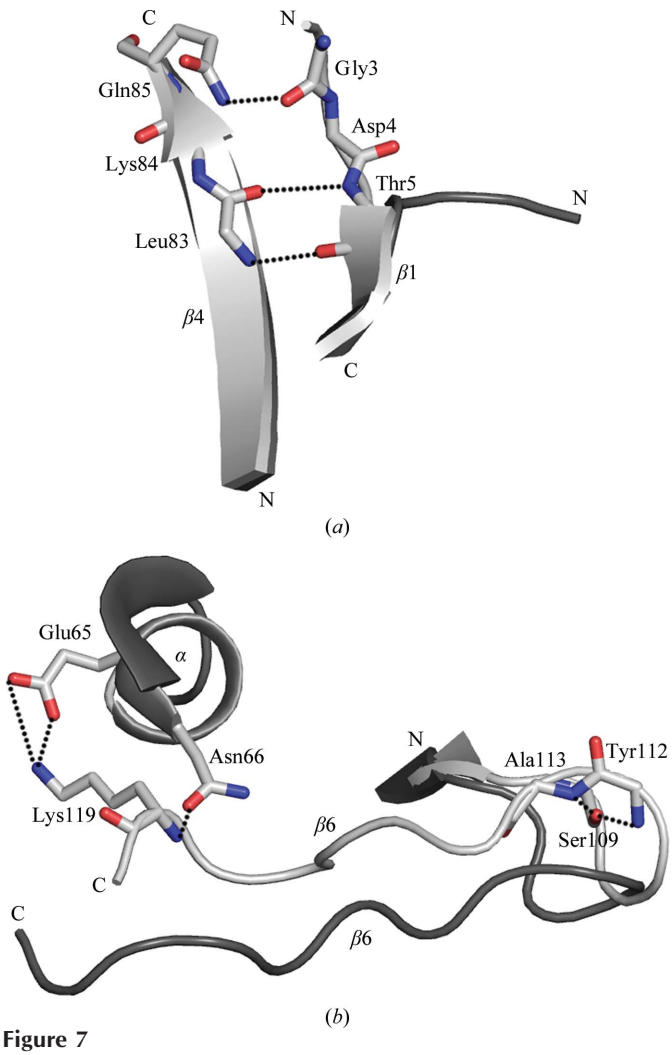

Figure 7
Conformational differences between the MlSSB tetramer (black) and the individual subunit (light shade) for (*a*) the N-terminal stretch and (*b*) the 113–120 stretch. Only side chains involved in interactions are shown.

3.6. Dynamics of inter-subunit hydrogen bonds/water bridges

An interesting feature, to which special attention was paid, was the hydrogen bonds/water bridges between subunits *A* and *B* (and *C* and *D*). In this respect, the results of simulations involving the tetramers derived from EcSSB and MlSSB form I were unambiguous. In EcSSB, all the interactions in the four pairs remain as hydrogen bonds as in the starting structure. The distances indicate that three symmetric pairs of interactions remain water bridges in MlSSB as in the crystal structure. Water molecules appear in a sphere of 1.6 Å radius around the positions in the crystal structure, with occupancies in the range 0.2–0.5. The two outermost interactions appear to have disappeared, with the interacting atoms moving apart. However, the simulations involving the tetramer derived from form II MlSSB present a more complex picture. The direct hydrogen bonds in the *CD* interface remain as they are during the simulation, as do the outer four hydrogen bonds in the *AB*

interface. The time-evolution indicates that the inner interactions stabilize as water bridges, although they do access a situation involving direct hydrogen bonds in a transient manner. However, the tetramer settles in a situation similar to that in the crystal structure.

4. Conclusions

Tetrameric SSB proteins of known structure exhibit substantial variability in quaternary association, although all of them have essentially the same tertiary structure. The mycobacterial SSB proteins and ScSSB have two dimers in the tetramer that are clamped together through the swapping of a C-terminal strand which only exists in these SSB proteins. The variability in the mutual orientation of the monomers in the dimer is lower than that in the mutual orientation of the two dimers in the tetramer. Dimerization in some of the crystal structures of mycobacterial SSB proteins and that of ScSSB involves eight water bridges between twofold-related strands, while in some other structures these bridges are replaced by direct hydrogen bonds. Direct hydrogen bonds exist in all other SSB proteins. Form II MlSSB presents an interesting case in which in one of the dimers four water bridges in the middle are flanked by two weak hydrogen bonds on either side. The variability in quaternary structure does not wholly correlate with differences in amino-acid sequences, but it probably does correlate with protein stability as evidenced by the surface area buried on oligomerization.

Molecular-dynamics simulations together with crystal structures provide valuable insights into the relatively rigid and flexible regions of the subunits in MlSSB and EcSSB. They also shed light on the changes brought about in the subunits on account of oligomerization. The simulations indicate that in the free subunits the N-terminal stretch is flexible and could move away from the body of the protomer in EcSSB, while both terminal stretches remain anchored to the body of the MlSSB subunit even when it is not part of an oligomer. Interestingly, this differential behaviour can be explained on the basis of a single substitution in the amino-acid sequence. The C-terminal β-strand in MlSSB involved in clamping the two dimers together in the tetramer remains close to the body of the subunit even on dissociation of the oligomer on account of alternative hydrogen-bonding possibilities. During simulations, the direct hydrogen bonds between the monomers in the dimer of EcSSB remain intact, as do the six inner water bridges in the tetramer derived from form I MlSSB. The atoms involved in the two outer water bridges drift apart. The mixed situation in the tetramer derived from form II MlSSB remains stable, although it can access the all-direct hydrogen-bond situation. Dimers involving all inter-subunit hydrogen bonds and involving all water bridges would appear to represent two states which do not easily interconvert. An intermediate state could also remain dynamically stable.

Data were collected at the X-ray Facility for Structural Biology at the Molecular Biophysics Unit supported by the Department of Science and Technology, Government of India. Computations were performed partially at the Bioinformatics Centre and the Graphics Facility, both of which are supported by the Department of Biotechnology (DBT). The work forms part of a DBT-sponsored research programme. During the period of this work MV was supported successively by a DBT Distinguished Research Professorship and a DAE Homi Bhabha Professorship.

References

Bernstein, D. A., Eggington, J. M., Killoran, M. P., Misic, A. M., Cox, M. M. & Keck, J. L. (2004). *Proc. Natl Acad. Sci. USA*, **101**, 8575–8580.
Bradford, M. M. (1976). *Anal. Biochem.* **72**, 248–254.
Brünger, 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.
Chan, K.-W., Lee, Y.-J., Wang, C.-H., Huang, H. & Sun, Y.-J. (2009). *J. Mol. Biol.* **388**, 508–519.
Chetnani, B., Kumar, P., Surolia, A. & Vijayan, M. (2010). *J. Mol. Biol.* **400**, 171–185.
Cohen, G. E. (1997). *J. Appl. Cryst.* **30**, 1160–1161.
Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.
Darden, T. A., York, D. M. & Pedersen, L. G. (1993). *J. Chem. Phys.* **98**, 10089–10092.
Das, D., Hyun, H., Lou, Y., Yokota, H., Kim, R. & Kim, S.-H. (2007). *Proteins*, **67**, 776–782.
DeLano, W. L. (2002). *PyMOL.* http://www.pymol.org.
DiDonato, M. *et al.* (2006). *Proteins*, **63**, 256–260.
Emsley, P. & Cowtan, K. (2004). *Acta Cryst.* D**60**, 2126–2132.
Handa, P., Acharya, N., Thanedar, S., Purnapatre, K. & Varshney, U. (2000). *Nucleic Acids Res.* **28**, 3823–3829.
Hess, B., Bekker, H., Berendsen, H. J. C. & Fraaije, J. G. E. M. (1997). *J. Comput. Chem.* **18**, 1463–1472.
Hubbard, S. J. & Thornton, J. M. (1993). *NACCESS Computer Program.* Department of Biochemistry and Molecular Biology, University College London.
Jędrzejczak, R., Dauter, M., Dauter, Z., Olszewski, M., Długołęcka, A. & Kur, J. (2006). *Acta Cryst.* D**62**, 1407–1412.
Jorgensen, W. L. & Maxwell, D. S. (1996). *J. Am. Chem. Soc.* **118**, 11225–11236.
Kaushal, P. S., Talawar, R. K., Krishna, P. D. V., Varshney, U. & Vijayan, M. (2008). *Acta Cryst.* D**64**, 551–560.
Krishna, R., Manjunath, G. P., Kumar, P., Surolia, A., Chandra, N. R., Muniyappa, K. & Vijayan, M. (2006). *Nucleic Acids Res.* **34**, 2186–2195.
Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam, H., Valentin, F., Wallace, I. M., Wilm, A., Lopez, R., Thompson, J. D., Gibson, T. J. & Higgins, D. G. (2007). *Bioinformatics*, **23**, 2947–2948.
Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). *J. Appl. Cryst.* **26**, 283–291.
Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.
McDonald, I. K. & Thornton, J. M. (1994). *J. Mol. Biol.* **238**, 777–793.
Murshudov, G. N., Vagin, A. A. & Dodson, E. J. (1997). *Acta Cryst.* D**53**, 240–255.
Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
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.
Reddy, M. S., Guhan, N. & Muniyappa, K. (2001). *J. Biol. Chem.* **276**, 45959–45968.

Saikrishnan, K., Jeyakanthan, J., Venkatesh, J., Acharya, N., Sekar, K., Varshney, U. & Vijayan, M. (2003). *J. Mol. Biol.* **331**, 385–393.
Saikrishnan, K., Manjunath, G. P., Singh, P., Jeyakanthan, J., Dauter, Z., Sekar, K., Muniyappa, K. & Vijayan, M. (2005). *Acta Cryst.* D**61**, 1140–1148.
Schneider, T. R. (2002). *Acta Cryst.* D**58**, 195–208.
Selvaraj, M., Roy, S., Singh, N. S., Sangeetha, R., Varshney, U. & Vijayan, M. (2007). *J. Mol. Biol.* **372**, 186–193.
Shereda, R. D., Kozlov, A. G., Lohman, T. M., Cox, M. M. & Keck, J. L. (2008). *Crit. Rev. Biochem. Mol. Biol.* **43**, 289–318.
Spoel, D. van der, Lindahl, E., Hess, B., Groenhof, G., Mark, A. E. & Berendsen, H. J. (2005). *J. Comput. Chem.* **26**, 1701–1718.
Štefanić, Z., Vujaklija, D. & Luić, M. (2009). *Acta Cryst.* D**65**, 974–979.
Storoni, L. C., McCoy, A. J. & Read, R. J. (2004). *Acta Cryst.* D**60**, 432–438.
Vijayan, M. (2005). *Tuberculosis*, **85**, 357–366.
Winn, M. D., Murshudov, G. N. & Papiz, M. Z. (2003). *Methods Enzymol.* **374**, 300–321.
Yang, C., Curth, U., Urbanke, C. & Kang, C. (1997). *Nature Struct. Biol.* **4**, 153–157.