Structure of *Mycobacterium smegmatis* single-stranded DNA-binding protein and a comparative study involving homologus SSBs: biological implications of structural plasticity and variability in quaternary association

K. Saikrishnan,[a] G. P. Manjunath,[b] Pawan Singh,[b] J. Jeyakanthan,[a] Z. Dauter,[c] K. Sekar,[d] K. Muniyappa[b] and M. Vijayan[a]*

[a]Molecular Biophysics Unit, Indian Institute of Science, Bangalore 560 012, India, [b]Department of Biochemistry, Indian Institute of Science, Bangalore 560 012, India, [c]Synchrotron Radiation Research Section, MCL, National Cancer Institute, Brookhaven National Laboratory, Upton, NY 11973, USA, and [d]Bioinformatics Centre, Indian Institute of Science, Bangalore 560 012, India

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

The structure of *Mycobacterium smegmatis* single-stranded DNA-binding protein (SSB) has been determined using three data sets collected from related crystals. The structure is similar to that of its homologue from *Mycobacterium tuberculosis*, indicating that the clamp arrangement that stabilizes the dimer and the ellipsoidal shape of the tetramer are characteristic features of mycobacterial SSBs. The central OB fold is conserved in mycobacterial SSBs as well as those from *Escherichia coli*, *Deinococcus radiodurans* and human mitochondria. However, the quaternary structure exhibits considerable variability. The observed plasticity of the subunit is related to this variability. The crystal structures and modelling provide a rationale for the variability. The strand involved in the clamp mechanism, which leads to higher stability of the tetramer, appears to occur in all high-G+C Gram-positive bacteria. The higher stability is perhaps required by these organisms. The mode of DNA binding of mycobacterial SSBs is different from that of *E. coli* SSB partly on account of the difference in the shape of the tetramers. Another difference between the two modes is that the former contains additional ionic interactions and is more susceptible to salt concentration.

PDB References: low-temperature MsSSB using synchrotron radiation, 1x3e; low-temperature MsSSB using home source, 1x3f; room-temperature MsSSB using home source, 1x3g.

1. Introduction

The continued prevalence of tuberculosis (TB) is primarily a consequence of the ability of *Mycobacterium tuberculosis* to persist for decades in the host under highly unstable environmental conditions (Dye *et al.*, 1999; Stewart *et al.*, 2003). The difficulties involved in working with the tubercle bacillus have to some extent impeded the study of the molecular biology of this pathogen. The recognition that *M. smegmatis* can be used as a surrogate model represents a landmark in the study of the molecular genetics of *M. tuberculosis* (Jacobs *et al.*, 1991). As part of a concerted international effort (Terwilliger *et al.*, 2003), we have been involved in elucidating the structures of proteins from *M. tuberculosis* (Datta *et al.*, 2000; Saikrishnan *et al.*, 2003, 2004). We have also been augmenting our findings with studies of proteins from *M. smegmatis* (Datta *et al.*, 2003; Roy *et al.*, 2004). As part of this programme, the structure of *M. tuberculosis* single-stranded DNA-binding protein (MtSSB) has recently been solved (Saikrishnan *et al.*, 2003). This study revealed the protein to possess certain unique features in comparison to homologues from other sources,

with implications for the survival of the pathogen under highly unstable environmental conditions.

Single-stranded DNA-binding proteins (SSBs) are thought to protect the transient single-stranded DNA (ssDNA) generated during DNA metabolism from chemical and nuclease attacks and to prevent them from forming aberrant secondary structures. The key role played by SSBs in the maintenance of genomic integrity makes them one of the essential gene products required for growth and survival (Mushegian & Koonin, 1996). The structures of SSBs generally have a conserved folding domain, referred to as the oligonucleotide-binding (OB) fold (Suck, 1997). However, structural differences do arise, particularly in the form of variations in the multimeric state adopted by the SSBs. For example, the SSB from the phage T4, the gene 32 protein, exists as a monomer in solution, while the homologues from eukaryotes exist as heterotrimers (Suck, 1997). Bacterial SSBs and mitochondrial SSBs are predominantly homotetramers in solution. Among the homotetrameric SSBs, diversity is introduced by variation in the quaternary architecture (Saikrishnan *et al.*, 2003).

While SSBs from *Escherichia coli* (EcSSB; Raghunathan *et al.*, 1997; Webster *et al.*, 1997; Matsumoto *et al.*, 2000) and human mitochondria (HMtSSB; Yang *et al.*, 1997) have similar quaternary structures, MtSSB possesses a different tetrameric structure. A monomer of eubacterial SSB can be divided into two domains: the N-terminal DNA-binding domain and the C-terminal glycine/proline-rich tail, which mediates protein–protein interactions. A number of biochemical studies have been carried out on the prototypical EcSSB (Lohman & Ferrari, 1994) and the mycobacterial SSBs MtSSB and MsSSB (Purnapatre & Varshney, 1999; Reddy *et al.*, 2001; Acharya & Varshney, 2002). These studies indicate that despite substantial differences in the amino-acid sequences, the DNA-binding affinities displayed by EcSSB and mycobacterial SSBs are similar. Studies on mycobacterial repair and recombination apparatus revealed that MtSSB and MsSSB stimulate cognate RecA (Reddy *et al.*, 2001; Ganesh & Muniyappa, 2003*a*,*b*) and UDG (Acharya & Varshney, 2002). This stimulation has been suggested to involve the C-terminal domain. The possibility of the involvement of the N-terminal domain in such interactions has also been proposed (Saikrishnan *et al.*, 2003; Handa *et al.*, 2001).

Interestingly, studies on mycobacterial RecA revealed that while MtRecA formed a stable complex with MtSSB, MsSSB displayed low affinity for the cognate MsRecA as well as for the non-cognate MtRecA (Reddy *et al.*, 2001). This observation indicated the possibility of structural differences between the closely related MtSSB and MsSSB, although they share a sequence identity of 84%. We therefore sought to determine the X-ray crystal structure of MsSSB and compare its structure with that of MtSSB. We also present a comparative study of the mycobacterial SSBs with known tetrameric or four OB-fold domain SSB structures. In addition, this study provides structural insights into the plasticity of the protein molecule and the variability of its quaternary association. The biological implications of the results are discussed.

2. Materials and methods

2.1. Crystallization and data collection

MsSSB 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 precipitant (1 *M* sodium acetate, 500 m*M* sodium chloride and 50 m*M* cadmium sulfate in 20 m*M* Tris–HCl buffer pH 7.4). A room-temperature data set and a low-temperature data set were collected with a MAR Research imaging-plate detector mounted on a Rigaku X-ray generator to resolutions of 3 and 2.7 Å, respectively. Another low-temperature data set (2.15 Å resolution) was collected on the X9B beamline, NSLS, Brookhaven National Laboratory using an ADSC Quantum 4 CCD detector. The low-temperature data sets were collected at 100 K using 30% glycerol as the cryoprotectant. The *HKL* package was used for data processing and scaling (Otwinowski & Minor, 1997). *TRUNCATE* (Collaborative Computational Project, Number 4, 1994) was used to convert intensities to structure factors.

2.2. Structure determination and refinement

The structure was determined using the molecular-replacement program *AMoRe* (Navaza, 1994) for the three data sets independently. The solutions thus obtained were refined against the data sets in a similar manner using *CNS* v.1.1 (Brünger *et al.*, 1998) in the early stages. Iterations of rigid-body and positional refinement and simulated annealing were alternated with model building using *FRODO* (Jones, 1978). Water molecules were located based on peaks with height greater than 2.5σ in the $F_o - F_c$ maps and those with height greater than 0.8σ in $2F_o - F_c$ maps. Cadmium ions were defined on the basis of very strong electron density in the $2F_o - F_c$ maps. Bulk-solvent corrections and overall anisotropic *B*-factor corrections were used throughout the refinement. The refinement of the three models was completed using *REFMAC*5 (Murshudov *et al.*, 1996). In the case of the 2.15 Å structure, refinement of the temperature factors using TLS (Winn *et al.*, 2003) was carried out. Each subunit was divided into four TLS groups, with group I consisting of residues 2–21, 28–34, 53–85 and 98–120, group II consisting of residues 22–27, group III consisting of residues 35–52 and group IV consisting of residues 86–97. Group I represents the molecular core, while groups II, III and IV represent loops. The data and the refinement statistics are presented in Table 1. An omit electron-density map corresponding to β-strands 8 and 9 is shown in Fig. 1.

2.3. Structure modelling

Selected models of tetrameric SSBs were minimized using the following protocol. The models were soaked in a 5 Å shell of water using *INSIGHT*II after the H atoms had been generated. The models were subjected to energy minimization and simulated annealing using *CNS* v.1.1 (Brünger *et al.*, 1998). A dielectric constant of unity was used throughout. A main-chain restraint of 4.2 kJ mol^{-1} was applied to the protein

molecule. In the first step, the models were subjected to conjugate-gradient energy minimization with a small repulsive van der Waals term introduced and the electrostatic term switched off. In the next step, the electrostatic term was switched on and the structures minimized for 100 cycles each. Subsequently, the simulated-annealing protocol was used to remove ambiguities about the preferences of side-chain and main-chain torsions among the available rotamers. The models were heated to 3000 K and the simulations were performed in steps of 25 K, with each step containing 50 cycles spanning 5 fs each. Following simulated annealing, one more step of conjugate-gradient minimization was carried out until the gradient of the total energy converged or was below 4.2 kJ mol^{-1} Å^{-1}. A non-crystallographic restraint of 1 kJ mol^{-1} was imposed in all the above steps. Tethering restraints of 2.1 and 0.4 kJ mol^{-1} were applied to the main-chain and side-chain atoms, respectively. *CNS* v.1.1 (Brünger *et al.*, 1998) was used to calculate the interaction energy between the two polypeptide chains of the models, employing the distance-dependent dielectric constant.

Table 1
Data and refinement statistics.

Values in parentheses are for the final shell.

	Room temperature (home source) (298 K)	Low temperature (home source) (100 K)	Low temperature (synchrotron) (100 K)
Space group	$P3_121$	$P3_121$	$P3_121$
Unit-cell parameters (Å)			
a	79.9	78.0	78.6
b	79.9	78.0	78.6
c	81.3	71.0	79.8
Data resolution (Å)	3.0	2.7	2.15
Total No. of reflections	26083	45835	66544
Unique reflections	6048	7150	15708
Completeness (%)	96.1 (98.0)	99.4 (100.0)	98.6 (99.2)
R_{merge}	11.4 (48.1)	10.9 (52.6)	5.8 (49.2)
R_{meas}†	12.9 (54.5)	11.8 (57.6)	6.6 (56.1)
Refinement			
R factor	20.1	19.7	20.0
R_{free}	25.4	27.7	23.8
Resolution range (Å)	20.0–3.0	20.0–2.7	30.0–2.15
R.m.s. deviation from ideality			
Bonds (Å)	0.022	0.017	0.014
Angles (°)	1.9	1.6	1.4
Residues in Ramachandran plot‡ (%)			
Most favoured region	81.6	85.2	90.9
Additionally allowed region	12.8	11.1	6.6
Generously allowed region	3.6	3.7	2.5
Disallowed region	2.5	0.0	0.0
No. of protein atoms	1648	1631	1697
No. of water molecules	100	189	261
No. of cadmium ions	1	2	1

† As defined by Diederichs & Karplus (1997). ‡ Calculated for non-glycine and non-proline residues using *PROCHECK* (Laskowski *et al.*, 1993).

2.4. Geometrical analyses of structures

Structural superposition of C$^{\alpha}$ atoms was carried out and the r.m.s. deviation calculated using *LSQKAB* (Collaborative Computational Project, Number 4, 1994) and *ALIGN* (Cohen, 1997). Accessible surface areas were calculated using *NACCSESS* (Hubbard, 1996). A probe radius of 1.4 Å was used for this purpose. Surface complementarity was calculated using the method of Lawrence & Colman (1993) as implemented in the *CCP*4 suite of programs. Information from the structural superposition was used to generate a structure-based amino-acid sequence alignment of tetrameric SSBs of known structure. The plasticity of the molecule was determined using the program *ESCET* (Schneider, 2002). The molecule was delineated into rigid and flexible regions with a cutoff value of 2.3σ. The cutoff parameter involving σ was chosen so that the molecule was divided equally into rigid and flexible regions when representative structures from MsSSB, MtSSB, EcSSB, HMtSSB and DrSSB were used. The two subunits in the asymmetric unit of the 2.15 Å structure of MsSSB, two and four subunits from the form I (PDB code 1ue1) and form II (PDB code 1ue6) crystal structures of MtSSB, respectively, one subunit from free EcSSB (PDB code 1kaw), four subunits of DNA-bound EcSSB (PDB code 1eyg), two subunits of HMtSSB (PDB code 1s3o) and the two OB-fold domains (residues 4–109 and 128–231) from a subunit of DrSSB (PDB code 1se8) were used in the above calculations. Owing to the large variability in the structure of the three

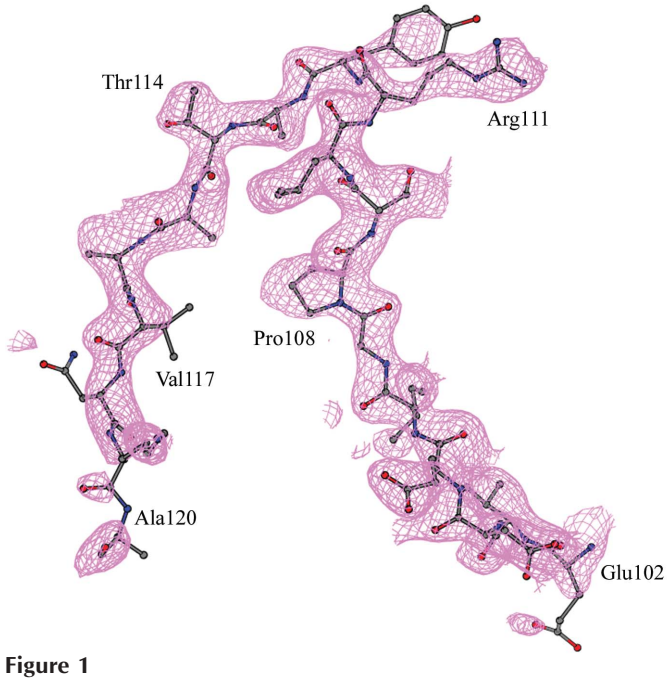

Figure 1
A $2F_o - F_c$ omit map for the stretch of residues 102–120. The figure was generated using *BOBSCRIPT* (Esnouf, 1997). The map was contoured at 1σ.

loops and because of their incomplete definition in many of the structures, the plasticity of the molecules was calculated by excluding the loops.

3. Results and discussion

3.1. Models resulting from three data sets

The three data sets have somewhat different unit-cell parameters (Table 1). The room-temperature data give a unit-cell volume that is larger by 4.7% than the low-temperature data collected using synchrotron radiation. The variation is in line with those documented for other flash-frozen protein crystals (Juers & Matthews, 2001). Interestingly, the low-temperature data set collected using the home source has a unit-cell volume that is smaller by 12.4% than that collected using synchrotron radiation. The large difference in the volume of the two data sets collected from flash-frozen crystals, albeit using two different sources, is unusual. One possible cause for this variation is the difference in the treatment of the crystals with the cryoprotectant. The crystal used to collect the low-temperature data obtained using the home source was soaked in glycerol by raising the concentration of glycerol stepwise from 0 to 30% over a period of 6 h and then flash-freezing immediately in a stream of liquid nitrogen. On the other hand, the crystal that was exposed to synchrotron radiation was, subsequent to the initial stepwise treatment with glycerol, preserved in mother liquor containing 30% glycerol for almost half a month before flash-freezing.

The structures of MsSSB refined against the three data sets do not display appreciable structural variation despite significant differences in the unit-cell volume. The r.m.s. deviations in C^{α} positions in the structures refined against the room-temperature and the low-temperature data with respect to that refined against the synchrotron data are 0.76 and 0.80 Å, respectively. The change in the unit-cell volume is instead accommodated by a change in packing density. While each molecule in the asymmetric unit of the synchrotron structure and the room-temperature structure make 170 and 140 contacts at 3.6 Å cutoff, respectively, with their neighbours, the low-temperature structure (home source), which has the smallest unit-cell volume, makes 256 contacts, indicating significantly tighter molecular packing in this case.

3.2. Molecular architecture

Although the full-length protein was crystallized, the electron density for residues beyond 120 did not exist, as in other bacterial SSB structures. The 120 N-terminal residues constitute the DNA-binding domain. The structure of this domain is very similar to that in the trigonal crystal form (form I) of MtSSB. The few differences between the two SSBs in the sequence of this domain do not affect the structure. This indicates that the N-terminal domain does not contribute to the difference between the behaviour of MsSSB and MtSSB towards RecA. The difference is possibly caused by changes in the sequence of the C-terminal domain. The C-terminal tail of MsSSB appears to be more flexible on account of its higher glycine content (Reddy *et al.*, 2001).

The C-terminal tail domain is common to bacterial SSBs, but has so far eluded structural characterization. In this context, it is interesting to note that a number of protein and protein domains have been found to exist in unfolded states (Wright & Dyson, 1999). Natively unfolded proteins are characterized by a very low frequency of hydrophobic amino acids and a higher net charge (Uversky *et al.*, 2000). The 117–165 stretch of residues in MsSSB was identified as disordered by *FoldIndex* (Zeev-Ben-Mordehai *et al.*, 2003; http://bip.weizmann.ac.il/miwbin/servers), which predicts intrinsically disordered protein based on the above idea. A similar observation has been made in the case of the full-length EcSSB crystal structure (Savvides *et al.*, 2004). As has been suggested in the case of the natively unfolded protein gliotactin (Zeev-Ben-Mordehai *et al.*, 2003), it appears that the extended length, high flexibility and inherent plasticity lend the C-terminal domain of SSB the ability to efficiently recruit diverse sets of proteins involved in DNA transactions to the site of action. It is also possible that the C-terminal domain becomes structured upon interacting with its protein partners.

The 2.15 Å structure of MsSSB is derived from one of the best resolved data sets for bacterial SSBs and provides the most accurate molecular model for this protein from mycobacteria. The polypeptide chain adopts the OB fold (Murzin, 1993). Its structure is characterized by three long β-hairpin loops (residues 22–27, 36–52 and 85–98) extending out of a globular core formed by a five-stranded β-barrel, which is capped by an α-helix (Fig. 2). As in MtSSB, the C-terminus of

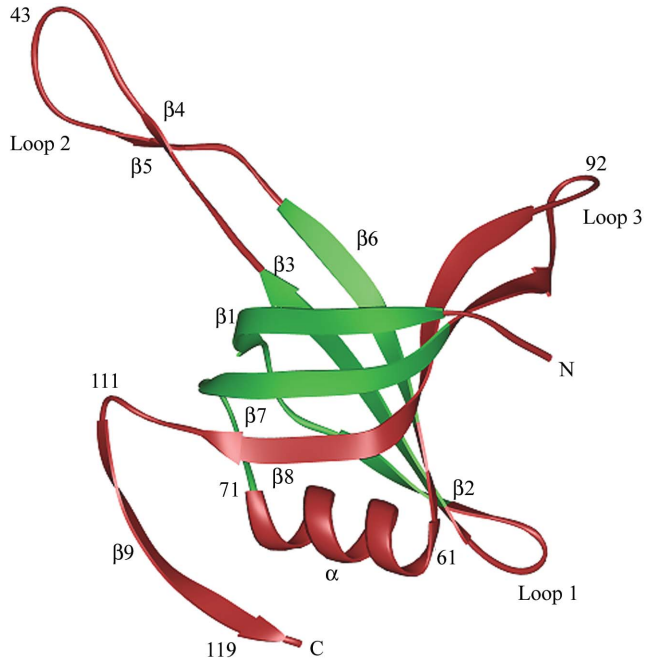

Figure 2
Molecular structure of MsSSB. The relatively rigid and flexible regions of the SSB molecule mapped onto the MsSSB structure are shown in green and red, respectively. The figure was generated using the program *RIBBONS* (Carson, 1997).

the barrel makes a turn and extends as a β-strand (β-strand 9) forming a hook-like structure along with the preceding strand. This strand is a unique feature of mycobacterial SSBs and has not been observed in any other SSBs studied to date. The architecture of the MsSSB tetramer (Fig. 3), which is a dimer of dimers, is similar to that in the form I crystal of MtSSB (Saikrishnan *et al.*, 2003). The dimeric interface is generated by the back-to-back association of the three-stranded back β-sheet and a clamping mechanism involving strand 9. The dimers associate to form the tetramer. This association involves the side-by-side arrangement of the back β-sheets.

3.3. Loop motion in MsSSB

During the refinement of the structure, it was noted that the temperature factors of the three loops were considerably higher than that of the molecular core, implying greater structural flexibility. The inherent flexibility of the loops appears to be common to most other SSBs. Consequently, in most of the structures of bacterial SSBs the loops are modelled incompletely owing to ill-defined electron density. Even in the recently determined 1.8 Å structure of *Deinococcus radiodurans* SSB (DrSSB) certain portions of the loops have not been defined (Bernstein *et al.*, 2004). In the three structures of MsSSB reported here, the loops have been defined to varying extents. All the residues of loop 1 have been located in all the three structures. The other two loops have been defined completely in at least one of the two subunits in the asymmetric unit of each of the three structures. The higher resolution of one of the MsSSB X-ray diffraction data sets afforded the refinement of the temperature factors of the structure using translation–libration–screw (TLS) parameters (Winn *et al.*, 2003), as elaborated in §2. As part of the TLS refinement, the three loops in each subunit of the asymmetric unit were considered as independent rigid bodies. The plot of the temperature factors along the polypeptide chain of the 2.15 Å structure (Fig. 4) obtained subsequent to the TLS refinement indicates that the higher temperature factor in loop 2 and 3 result from rigid-body motion.

A comparison of the loop orientation in different crystallographically independent subunits revealed that in addition to conformational changes within the loops, as seen from the high r.m.s. deviations for these loops, they also demonstrate rigid-body movement. The motion in all the cases is essentially rotation. The short loop 1 is relatively less mobile, with an angular displacement in the range of 10°. Loop 2, the longest of the three loops, shows angular displacement in the range of 20°. However, in both cases the orientation adopted by the loops in the subunits appears to be random, as the axes of rotation that relate the various orientations are very different. Among the three loops, loop 3 displays the largest movement. For example, the loop in subunits *A* and *B* of the 2.15 Å structure is oriented so differently that a rotation of 47° is required to superpose them (Fig. 5). In situations where the angles of rotation are large, the axes of rotation required to superpose loop 3 are nearly parallel. The difference in orientation of loop 3 in the two subunits of the asymmetric unit appears to be the result of crystal packing. A neighbouring symmetry-related molecule docks onto the DNA-binding cleft of subunit *B*, which results in the widening of this cleft and hence the movement of loop 3 of this subunit. Interestingly, the axes of rotation and the libration axis obtained from the TLS refinement of loop 3 of subunit *A* of the 2.15 Å structure are roughly parallel (angle < 22°) (Fig. 5). Thus, the motion necessary to accommodate crystal packing

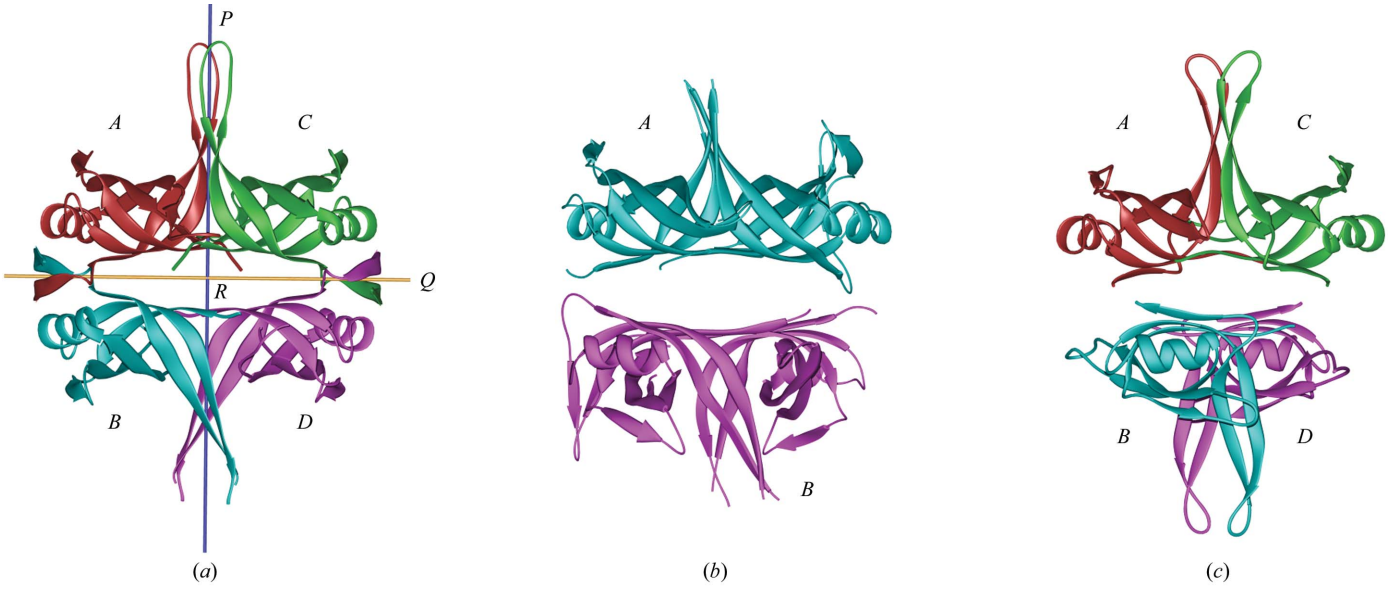

Figure 3
The quaternary structure of MsSSB (*a*) along with those of DrSSB (*b*) and EcSSB (c). While MsSSB and EcSSB are tetramers, DrSSB is a dimer, witheach monomer made up of two OB-fold domains. The three perpendicular twofold axes *P*, *Q* and *R*, which relate the subunits of the MsSSB tetramer, are depicted.

appears to have made use of the natural librational tendency of the loop.

3.4. Structural variations in tetrameric SSBs

MsSSB has a quaternary structure that is similar to that of MtSSB. The variations in quaternary structure among SSBs stem from differences in the orientation of the subunits that associate to form the two major interfaces of the dimer of dimers (Figs. 3 and 6). One of the major interfaces (that between *A* and *C* and between *B* and *D*), which is formed by the side-by-side arrangement of the subunits, is nearly the same in both mycobacterial and *E. coli* SSBs. The other major interface (that between *A* and *B* and between *C* and *D*), formed by back-to-back arrangement of the subunits, is different in the two cases. The quaternary arrangement in EcSSB can however be related to that in MsSSB by a rotation of 43° of two subunits (*B* and *D*) with respect to the other two (*A* and *C*). The axis of this rotation coincides with the twofold axis *P* in Fig. 3. In a dimer of dimers, other things being equal, the subunits that share the largest interface area can be designated as the dimer. On this basis, subunits *A* and *C*, which are arranged side-by-side, make the dimer in EcSSB, while subunits *A* and *B* make the dimer in MsSSB and MtSSB. This difference is consistent with the involvement of strand 9 in clamping the two subunits in the dimer of mycobacterial SSBs.

The structure of the bacterial SSB from *D. radiodurans* was reported recently with a variant quaternary structure (Bernstein *et al.*, 2004). DrSSB, unusually for bacterial SSB, is a homodimer. Each monomer of DrSSB is made up of two non-identical OB-fold domains fused together by a β-hairpin linker peptide. The arrangement of the OB folds in the monomer of DrSSB is similar to the side-by-side arrangement of subunits *A* and *C* (or *B* and *D*) in EcSSB and MsSSB (Fig. 3). The dimer is formed by back-to-back association of the subunits. Like other eubacterial SSBs, the functional unit of DrSSB is made up of four OB folds, which are arranged side-by-side and back-to-back (Bernstein *et al.*, 2004). On comparison with EcSSB and MsSSB, we find that the back-to-back arrangement of the OB folds in DrSSB is intermediate between the other two (Fig. 6). A gross arrangement of the OB folds in DrSSB can be obtained from EcSSB by rotating subunits *B* and *D* with respect to *A* and *C* by 25° and from MsSSB by a similar rotation of −21°.

Interestingly, it turns out that the plasticity of the DNA-binding domain itself is substantially related to the variability in oligomerization. By comparing the structures of the domain in different SSBs, the relatively invariant (rigid) and variable (flexible) regions of the domain were delineated (Fig. 2) using program *ESCET* (Schneider, 2002), as explained in §2. The three loops are intrinsically flexible. β-Strand 9, which is formally designated as flexible, occurs only in the mycobacterial proteins. This strand plays a crucial role in stabilizing the MsSSB and MtSSB dimers. In the β-barrel, only β-strand 8 is flexible. This strand is involved in back-to-back interactions in the MsSSB and MtSSB tetramers, but only a few residues in the strand are involved in such interactions in other SSBs. The helix that caps the barrel is also designated as flexible. The

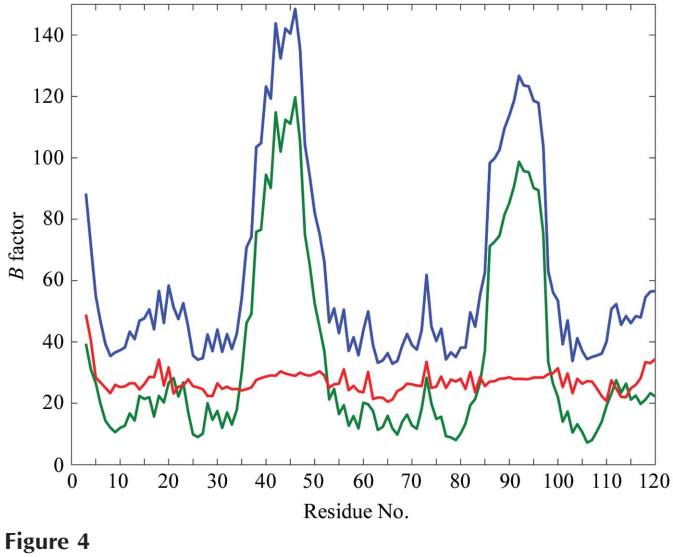

Figure 4
Plot of temperature factor along the polypeptide chain in subunit *B* of the 2.15 Å MsSSB structure. The total *B* factor (blue), the TLS component (green) and the residual *B* factor (red) are plotted.

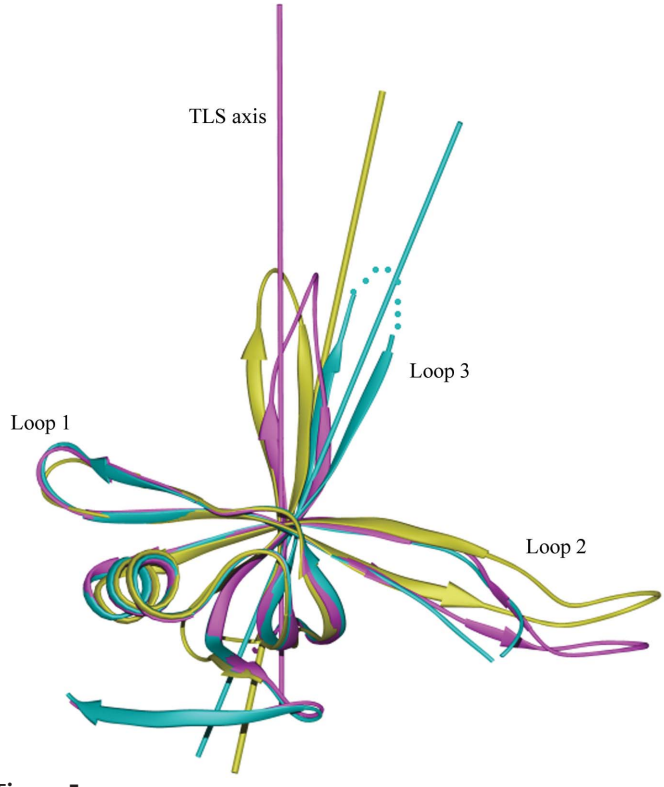

Figure 5
Structural superpositions of subunits *A* (magenta) and *B* (cyan) of MsSSB in the asymmetric unit of the 2.15 Å structure and a subunit of EcSSB from the crystal structure of the EcSSB–ssDNA complex (yellow). The orientation of subunit *B* of MsSSB and the subunit of EcSSB were obtained subsequent to the superposition of their respective cores onto the core of subunit *A* of MsSSB. Note the large deviation in the orientation of the loops. Also illustrated are the libration axis of loop 3 of subunit *A* (magenta) of MsSSB, the axis of rotation required to superpose loop 3 of subunit *B* of MsSSB (cyan) and the subunit of EcSSB (yellow) on to that of subunit *A* of MsSSB.

helix has somewhat different lengths in the two covalently linked domains in DrSSB. However, even when DrSSB is deleted from the calculations, parts of the helix remain flexible. The helix is again a region involved in quaternary interactions in the mycobacterial SSBs, but not in the other SSBs.

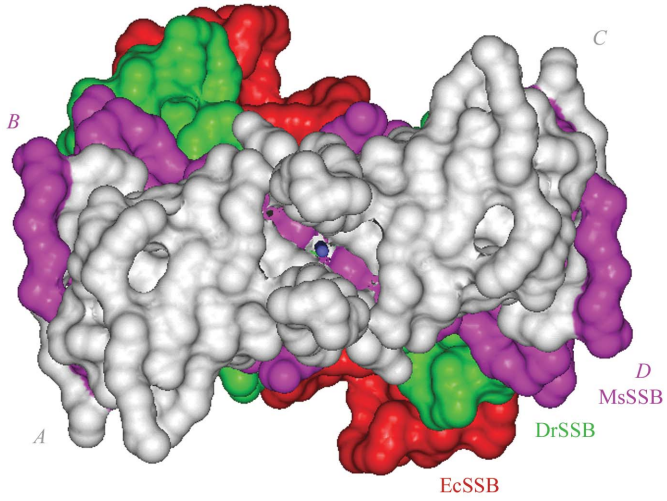

Figure 6
Surface diagram of MsSSB, DrSSB and EcSSB viewed down the twofold axis *P* (represented as a blue ball), highlighting the variation in the orientation of the OB folds. Subunits *A* and *C* of MsSSB and EcSSB and subunit *A* of DrSSB were superposed onto one another (represented in grey) to bring out the difference in the orientation of subunits *B* and *D* of MsSSB (magenta) and EcSSB (red) and subunit *B* of DrSSB (green). The loops are omitted for clarity.

3.5. Structural rationale for variation in quaternary structure

In EcSSB and HMtSSB, the back-to-back arrangement is locked in position by the formation of salt bridges across the interface involving residues Lys7 and Glu80, and Arg16 and Glu95, respectively (Raghunathan *et al.*, 1997; Yang *et al.*, 1997). In mycobacterial SSBs the presence of strand 9 at the C-terminus of the OB fold sterically prevents it from adopting the EcSSB or the DrSSB type of quaternary structure. Rearranging the subunits of mycobacterial SSBs to be similar to the arrangement in either EcSSB or DrSSB results in a severe steric clash between strand 9 of the subunits. As a consequence of the change in the interface geometry, the amino acids that populate the interfacial surfaces undergo suitable changes.

The cause for the differential arrangement of the OB folds in DrSSB is not immediately obvious. Neither the salt bridges found at the tetrameric interface in EcSSB and HMtSSB nor the clamp seen in mycobacterial SSBs are present in DrSSB. A simple-minded approach was adopted to explain the choice of the observed quaternary arrangement in DrSSB. The OB-fold domains in DrSSB were arranged like those in EcSSB and MsSSB by rotating subunit *B* by −25 and 21°, respectively, about the pseudo-twofold axis relating the two OB-fold domains. Both the arrangements were feasible, although they contained several steric clashes. The models were energy-minimized to remove short contacts. For comparison, the native structure of DrSSB was also minimized. The interaction energies between the two polypeptide chains, each consisting of two OB-fold domains, the surface area buried on their association and the surface complementarity between the two surfaces were calculated for all three models. The values of these parameters are given in Table 2. On every count, a quaternary arrangement for DrSSB similar to that of either EcSSB or mycobacterial SSB has parameters that are less favourable than those for the native arrangement. Thus, it would appear that it is the cumulative effect of overall changes rather than the effect of a few striking changes which is responsible for the distinctly different quaternary structure in DrSSB.

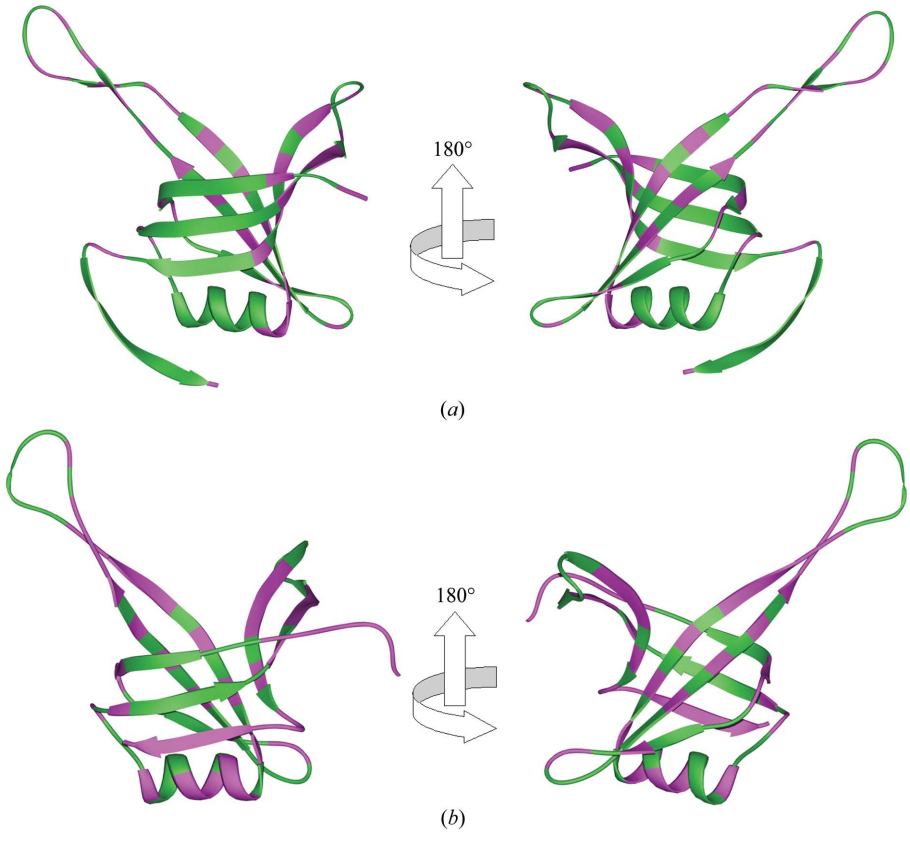

Figure 7
The monomeric structures of (*a*) MsSSB and (*b*) EcSSB depicting the residues, in magenta, interacting with ssDNA upon complexation.

3.6. Biological implications

An alignment of amino-acid sequences of the DNA-binding domain

Table 2
Surface area buried at the interface (Å^2), surface complementarity and interaction energy of DrSSB subunits in three different types of quaternary arrangement.

Arrangement	Total buried surface area	Hydrophobic buried surface area	Surface complementarity	Interaction energy (kJ mol^{-1})
DrSSB	2171	1469	0.66	−983
EcSSB	2024	1345	0.52	−858
MsSSB	1892	1225	0.62	−255

of SSBs from representative bacterial and mitochondrial species based on SSBs of known three-dimensional structure (data not shown) indicates that the critical β-strand 9 in mycobacterial SSBs is an insertion at the C-terminus of the OB fold. In fact, this insertion occurs in all high-G+C Gram-positive bacteria. A phylogenetic tree derived from the sequence alignment shows that the high-G+C Gram-positive bacteria cluster together. The tree indicates that the insertion is not of recent evolutionary origin but was possibly acquired by the ancestor of the present-day high-G+C Gram-positive bacteria. Thus, as in the case of mycobacterial SSBs, the quaternary association of the molecule is likely to be sturdy in all organisms with high-G+C genomes. The role of SSB in these organisms may be more critical and frequently required, as ssDNA with higher G+C content tends to form larger and more stable secondary structures.

Upon complexation, ssDNA wraps around the SSB oligomer (Raghunathan *et al.*, 2000). Owing to the variation in the quaternary structure, the DNA-binding surface in the case of EcSSB is an approximate spheroid, while that in the case of MsSSB is an ellipsoid. On account of the close similarity of MsSSB and MtSSB, the mode of DNA binding in the former is likely to be similar to that in the latter (Saikrishnan *et al.*, 2003). Fig. 7, which illustrates a map of the residues that line the DNA-binding site in the monomers of MsSSB and EcSSB, highlights the disparity that exists between the binding surface of MsSSB and that of EcSSB. The difference is most conspicuous around the helical region of the OB fold. In the EcSSB–DNA complex, ssDNA follows a shallow groove made up of hydrophobic and polar residues from the α-helix and the C-terminal strand (Raghunathan *et al.*, 2000). In mycobacterial SSBs, this groove is made inaccessible to DNA binding by the variant quaternary structure and the β-strand clamp, which together plug the groove. It turns out that many residues in the DNA-binding surface are not conserved in mycobacterial SSB and EcSSB. A larger number of positively charged ionic residues are observed on the DNA-binding surface of MsSSB in comparison to EcSSB. In concordance with this model, biochemical studies have indicated the stability of the mycobacterial SSB–DNA complex to be inversely related to the salt concentration, unlike that in the case of EcSSB (Reddy *et al.*, 2001).

Some of the residues constituting loop 3 are known to interact with ssDNA and stabilize the SSB–DNA complex (Raghunathan *et al.*, 2000). The crystal structure of EcSSB bound to ssDNA revealed a conformational change in loop 3 as a consequence of DNA binding, resulting in the movement of a residue, Trp88, by almost 2 Å (Raghunathan *et al.*, 2000). The orientations of this loop in DNA-bound EcSSB and MsSSB are related by a rotation of $\sim$26° (Fig. 5). Interestingly, the axis of rotation for this transformation is once again roughly parallel (angle of $<$33°) to the axis of libration displayed by loop 3 in MsSSB crystals (see §3.3). It would thus appear that loop 3 in mycobacterial SSB is designed to adopt a multiple set of conformations allowed by the invariant rotation axis and that upon binding to ssDNA this loop adopts one of the allowed conformations and closes onto the DNA.

X-ray intensity data were collected at the X-ray Facility for Structural Biology, supported by the Department of Science and Technology (DST) and Biotechnology (DBT), Government of India, and the X9B beamline, NSLS, Brookhaven National Laboratory, USA. Computations were performed at the Supercomputer Education and Research Centre at the Institute and the Bioinformatics Centre and the Interactive Graphics Facility, both funded by the DBT. The work forms part of a programme on Structural Genomics of Microbial Pathogens, supported by the DBT. MV is supported by a Distinguished Biotechnologist Award from the DBT.

References

Acharya, N. & Varshney, U. (2002). *J. Mol. Biol.* **318**, 1251–1264.
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.
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.
Carson, M. (1997). *Methods Enzymol.* **277**, 493–505.
Cohen, G. E. (1997). *J. Appl. Cryst.* **30**, 1160–1161.
Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.
Datta, S., Krishna, R., Ganesh, N., Chandra, N. R., Muniyappa, K. & Vijayan, M. (2003). *J. Bacteriol.* **185**, 3935–3947.
Datta, S., Prabu, M. M., Vaze, M. B., Ganesh, N., Chandra, N. R., Muniyappa, K. & Vijayan, M. (2000). *Nucleic Acid Res.* **28**, 4964–4973.
Diederichs, K. & Karplus, P. A. (1997). *Nature Struct. Biol.* **4**, 269–275.
Dye, C., Scheels, S., Dolin, P., Pathania, V. & Raviglione, M. C. (1999). *J. Am. Med. Assoc.* **282**, 677–686.
Esnouf, R. M. (1997). *J. Mol. Graph.* **15**, 132–134.
Ganesh, N. & Muniyappa, K. (2003*a*). *Biochemistry*, **42**, 7216–7225.
Ganesh, N. & Muniyappa, K. (2003*b*). *Proteins*, **53**, 6–17.
Handa, P., Acharya, N. & Varshney, U. (2001). *J. Biol. Chem.* **276**, 16992–16997.
Hubbard, S. J. (1996). *NACCESS* v.2.1.1. Department of Biomolecular Sciences, UMIST, Manchester, England.
Jacobs, W. R. Jr, Kalpana, G. V., Cirillo, J. D., Pascopella, L., Snapper, S. B., Udani, R. A., Jones, W., Barletta, R. G. & Bloom, B. R. (1991). *Methods Enzymol.* **204**, 537–555.
Jones, T. A. (1978). *J. Appl. Cryst.* **11**, 268–272.
Juers, D. H. & Matthews, B. W. (2001). *J. Mol. Biol.* **311**, 851–862.
Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). *J. Appl. Cryst.* **26**, 283–291.
Lawrence, M. C. & Colman, P. M. (1993). *J. Mol. Biol.* **221**, 327–346.
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. & Yasuoka, N. (2000). *J. Biochem.* **127**, 329–335.
Murshudov, G., Vagin, A. & Dodson, E. (1996). *Proceedings of the CCP4 Study Weekend. Macromolecular Refinement*, edited by K. S. Wilson, G. Davies, A. W. Ashton & S. Bailey, pp. 93–104. Warrington: Daresbury Laboratory.
Murzin, A. (1993). *EMBO J.* **12**, 861–867.
Mushegian, A. R. & Koonin, E. V. (1996). *Proc. Natl Acad. Sci. USA*, **93**, 10268–10273.
Navaza, J. (1994). *Acta Cryst.* A**50**, 157–163.
Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
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 (1997).
Reddy, M. S., Guhan, N. & Muniyappa, K. (2001). *J. Biol. Chem.* **276**, 45959–45968.
Roy, S., Gupta, S., Das, S., Sekar, K., Chatterji, D. & Vijayan, M. (2004). *J. Mol. Biol.* **339**, 1103–1113.
Saikrishnan, K., Jeyakanthan, J., Venkatesh, J., Acharya, N., Sekar, K., Varshney, U. & Vijayan, M. (2003). *J. Mol. Biol.* **331**, 385–393.
Saikrishnan, K., Kalapala, S. K., Bidya Sagar, M., Rao, A. R., Varshney, U. & Vijayan, M. (2004). *Acta Cryst.* D**60**, 368–370.
Savvides, S. N., Raghunathan, S., Fütterer, K., Kozlov, A. G., Lohman, T. M. & Waksman, G. (2004). *Protein Sci.* **13**, 1942–1947.
Schneider, T. R. (2002). *Acta Cryst.* D**58**, 195–208.
Stewart, G. R., Robertson, B. D. & Young, D. B. (2003). *Nature Rev. Microbiol.* **1**, 97–104.
Suck, D. (1997). *Nature Struct. Biol.* **4**, 161–165.
Terwilliger, T. C. *et al.* (2003). *Tuberculosis (Edinb.)*, **83**, 223–249.
Uversky, V. N., Gillespie, J. R. & Fink, A. L. (2000). *Proteins*, **41**, 415–427.
Webster, G., Genschel, J., Curth, U., Urbanke, C., Kang, C. & Hilgenfeld, R. (1997). *FEBS Lett.* **411**, 313–316.
Winn, M. D., Murshdov, G. N. & Papiz, M. Z. (2003). *Methods Enzymol.* **374**, 300–321.
Wright, P. E. & Dyson, H. J. (1999). *J. Mol. Biol.* **293**, 321–331.
Yang, C., Curth, U., Urbanke, C. & Kang, C. (1997). *Nature Struct. Biol.* **4**, 153–157.
Zeev-Ben-Mordehai, T., Rydberg, E. H., Solomon, A., Toker, L., Auld, V. J., Silman, I., Botti, S. & Sussman, J. L. (2003). *Proteins*, **53**, 758–767.