3D structure of *Thermus aquaticus* single-stranded DNA–binding protein gives insight into the functioning of SSB proteins

Roman Fedorov, Gregor Witte, Claus Urbanke, Dietmar J. Manstein and Ute Curth*

Institute for Biophysical Chemistry, Hannover Medical School, Carl-Neuberg-Strasse 1, D-30625 Hannover, Germany

Received July 28, 2006; Revised October 26, 2006; Accepted October 28, 2006

PDB accession codes: 2ihe and 2ihf

ABSTRACT

In contrast to the majority of tetrameric SSB proteins, the recently discovered SSB proteins from the *Thermus/Deinoccus* group form dimers. We solved the crystal structures of the SSB protein from *Thermus aquaticus* (TaqSSB) and a deletion mutant of the protein and show the structure of their ssDNA binding domains to be similar to the structure of tetrameric SSBs. Two conformations accompanied by proline cis-trans isomerization are observed in the flexible C-terminal region. For the first time, we were able to trace 6 out of 10 amino acids at the C-terminus of an SSB protein. This highly conserved region is essential for interaction with other proteins and we show it to adopt an extended conformation devoid of secondary structure. A model for binding this region to the χ subunit of DNA polymerase III is proposed. It explains at a molecular level the reason for the ssb113 phenotype observed in *Escherichia coli*.

INTRODUCTION

Bacterial single-stranded DNA–binding (SSB) proteins play an important role in DNA metabolism. They protect the vulnerable single-stranded DNA (ssDNA) from nucleolytic digestion and prevent hairpin formation, thus keeping ssDNA in a suitable conformation for the action of enzymes involved in processes like DNA replication, repair and recombination. Beyond this preparation of the ssDNA for proper enzymatic function, SSB proteins are involved in physical interactions with a couple of proteins that play an important role in those pathways (1–6). The presence of SSB proteins is essential for the survival of bacteria (7) and they are believed to belong to the minimal set of genes required for cellular life (8).

The primary structure of bacterial SSB proteins contains three distinct regions. The N-terminal DNA-binding domain, which forms an oligonucleotide/oligosaccharide-binding (OB)-fold (9), is succeeded by a glycine and proline-rich region and the acidic, highly conserved region formed by the last 10 amino acids. Almost all bacterial SSB proteins are active as tetramers where four OB-folds act together in the binding of ssDNA. The interaction of the four OB-folds seems to be essential for in vivo function, since a mutant SSB protein of *Escherichia coli* (EcoSSB) that only forms homodimers is not able to take over the function of EcoSSB in the *E.coli* cell (10).

Only recently dimeric SSB proteins have been discovered in the *Thermus/Deinoccus* group of bacteria (11–13). The protomers of these proteins are nearly twice the size of the tetrameric bacterial SSB proteins and contain two OB-folds per monomer (14). This observation suggests a gene duplication event in the evolution of the *Thermus/Deinoccus* group of bacteria (13). Therefore, the cooperation of four OB-folds in the function of bacterial SSB proteins seems to be a common principle.

While the structures of the DNA-binding domains of SSB proteins from *E.coli* (15–17), *Mycobacterium tuberculosis* (18), *Mycobacterium smegmatis* (19) and *Deinococcus radiodurans* (14) have been solved, there is little structural information about the C-terminal region of SSB proteins. Since the glycine- and proline-rich region of EcoSSB seems to be disordered (15,20), the highly conserved region formed by the last 10 amino acids could not be resolved in the crystals. However, it was shown that the negatively charged C-terminal region is essential for the survival of *E.coli* cells (21) and interacts with proteins involved in DNA replication, repair and recombination. Among these proteins are the χ subunit of the DNA polymerase III (4.5), exonuclease I (2), uracil DNA glycosylase (3) and priA helicase (1). In contrast, the glycine- and proline-rich region is not essential for the survival of *E.coli* cells and it was speculated to act as a spacer (21).
Here, we describe the crystal structures of wild-type SSB protein from the thermophilic, eubacterial organism, *Thermus aquaticus* (TaqSSB) and a mutant form of the protein that lacks the glycine- and proline-rich region (TaqSSBΔ228–252 R190A). In both structures the main part of the protein consists of an elongated β-barrel domain with two adjacent α-helices, which together form two OB-folds (9). For the wild-type protein we show that the glycine- and proline-rich C-terminal part of the protein can adopt two different conformations. In the structure of the deletion mutant, which lacks this region, we were able to trace 6 out of 10 C-terminal amino acids that are essential for protein–protein interactions. These residues are highly disordered and devoid of secondary structure. We present a model that shows the C-terminal region of SSB protein docked to the χ subunit of *E.coli* DNA polymerase III (22).

**MATERIALS AND METHODS**

**Buffers and reagents**

All materials were of the highest purity available and were obtained from Sigma and Fluka. *T.aquaticus* strain ATCC25104 was purchased from German Collection of Microorganisms and Cell Cultures (DSMZ).

**Cloning of the Thermus aquaticus ssb gene**

PCR amplification of the *ssb* gene was performed with heat disrupted *T.aquaticus* cells and primers (MWG Biotech) 5′- GGAGACCATGGCTAGGCTGCTAACC-3′ for forward priming and 5′-GAAGAAGATCTCTAAACACGAGCAAA-TCCTCTCCTA-3′ for reverse priming of the complete *ssb* gene. PCR was carried out using *Pfu* DNA polymerase (Stratagene), 200 μM dNTP and 0.5 μM of each primer with an annealing temperature of 54°C. After removal of nucleotides and proteins (PCR Purification Kit, Qiagen) the PCR product was digested with NcoI and BamHI and cloned into the NcoI and BamHI sites of plasmid pET15b (Novagen) to generate pETTaqSSB. All DNA constructs were verified by sequencing of both DNA strands of the gene (MWG Biotech).

**Construction of mutant TaqSSB proteins**

Site-directed mutagenesis was performed using the Quik-Change mutagenesis kit (Stratagene). For construction of the deletion mutant TaqSSBΔ228–252 a NotI restriction site was introduced at bp 675–681 and an NdeI site downstream of the Taqssb stop-codon generating pETTaqSSBentry. After digestion of this construct with NotI and NdeI a double-stranded DNA cassette (MWG Biotech) encoding TaqSSB amino acids 253–264 was introduced to generate pETTaqSSBΔ228–252. Thus TaqSSBΔ228–252 is a fusion protein in which amino acids 1–227 of TaqSSB are followed by the last 12 amino acids of TaqSSB. TaqSSBΔ228–252 R190A was constructed by site-directed mutagenesis of pETTaqSSBΔ228–252, exchanging the codon of arginine 190 by an alanine codon.

**Expression and purification of proteins**

As the TaqSSB gene contains codons rarely used in *E.coli*, protein expression of TaqSSB and TaqSSB mutants was performed in the *E.coli* strain Rosetta (DE3) pLysS (Novagen). A 900 mL overnight culture grown in standard Luria–Bertani (LB) medium in the presence of 100 μg/mL ampicillin, 30 μg/mL chloramphenicol and 1% glucose at 37°C was used to inoculate 10 L of LB medium in an aerated fermenter and the cells were grown at 37°C. An aliquot of 100 mL of the overnight culture was used for plasmid preparation and subsequent DNA sequencing to confirm the sequence of the expressed gene shortly before induction. After growing the cells to an OD600nm=1.0 protein expression was induced by addition of 1 mM IPTG and was continued for 5 h. After harvesting the cells were washed with buffer W [0.2 M NaCl, 50 mM Tris–HCl pH 8.3, 1 mM EDTA and 10% (w/v) sucrose], resuspended in buffer S [buffer W containing 15 mM spermidin] and stored at −80°C.

Wild-type TaqSSB, TaqSSBΔ228–252 and TaqSSBΔ228–252 R190A proteins were purified as follows. After thawing the cells, one tablet of ‘complete EDTA free’ protease inhibitor (Roche) per 50 mL, 0.04% sodium desoxycholat and 0.1 mM PMSF were added. Sonication and removal of cell debris by preparative ultracentrifugation was followed by precipitation of non-thermophilic proteins by heat treatment at 65°C for 4 min. Proteins in the supernatant of a second preparative ultracentrifugation step were precipitated at 4°C by a drop by drop addition of polyethyleneimine pH 6.9 up to a final concentration of 0.4%, incubated for 10 min and pelleted. The pellet was washed with TGE400 [400 mM NaCl, 50 mM Tris–HCl pH 8.3, 1 mM EDTA and 20% (v/v) glycerol] and extracted with TGE700 [700 mM NaCl, 50 mM Tris–HCl pH 8.3, 1 mM EDTA and 20% (v/v) glycerol] and insoluble material was removed by centrifugation. The clear supernatant was precipitated with 25% (w/v) (NH4)2SO4 at 4°C, the pellet was redissolved in buffer L [2 M NaCl, 50 mM Tris–HCl pH 7.5, 1 mM EDTA and 10% (v/v) glycerol] and loaded onto a BlueSepharose column (23) equilibrated in the same buffer. After washing the column with B1 (buffer L containing 50 mM urea), the column was eluted with buffer B2 (buffer L containing 5 M urea). Fractions were analyzed by SDS–PAGE (24) and TaqSSB fractions were pooled, diluted to A280nm = 0.1 mM IPTG and was continued for 5 h. After harvesting the cells were washed with buffer W [0.2 M NaCl, 50 mM Tris–HCl pH 8.3, 1 mM EDTA and 20% (v/v) glycerol] and insoluble material was removed by centrifugation. The clear supernatant was precipitated with 25% (w/v) (NH4)2SO4 at 4°C, the pellet was redissolved in buffer L [2 M NaCl, 50 mM Tris–HCl pH 7.5, 1 mM EDTA and 10% (v/v) glycerol] and loaded onto a BlueSepharose column (23) equilibrated in the same buffer. After washing the column with B1 (buffer L containing 50 mM urea), the column was eluted with buffer B2 (buffer L containing 5 M urea). Fractions were analyzed by SDS–PAGE and TaqSSB fractions were pooled, concentrated using Vivaspin concentrators (Vivasience), dialyzed against storage buffer [1 M NaCl, 20 mM Tris–HCl pH 7.6, 1 mM EDTA, 10% (v/v) glycerol] and stored at −20°C. The purified proteins were at least 98% pure as judged from Coomassie Brilliant blue stained SDS-gels.

**Crystallization**

Both wild-type and mutant protein crystals were grown at 20°C by vapor diffusion using the sitting drop geometry. For wild-type TaqSSB 1.5 μL of protein (3.8 mg/mL in 10 mM Tris–HCl pH 8.0) and reservoir solutions were mixed, the latter contained 100 mM Imidazol pH 7.4, 35% (v/v) 2-ethoxyethanol and 200 mM calcium acetate.
The structure of wild-type TaqSSB was determined by molecular replacement (26) using coordinates of SSB from *D. radiodurans* (14), pdb code: 1se8) as an initial model. Refinement was continued with CNS (27) and SHELX (28). The structure determination of TaqSSBA228–252 R190A was initiated by a round of rigid body refinement using the protein part of the wild-type TaqSSB coordinates and continued by the same building/refinement procedure. The final electron density maps are shown in Figures 1 and 3.

The final models display good stereochemistry (see Table 1). Structures were overlaid with the programme COOT and adjusted manually. The coordinates and structure factor amplitudes for wild-type TaqSSB and TaqSSBA228–252 R190A structures have been deposited in the Protein Data Bank (PDB ID codes 2ihe and 2ihf, respectively).

### Tracing flexible protein regions

To facilitate tracing of the glycine-rich C-terminal part of wild-type TaqSSB we have collected 100% complete high-redundancy diffraction data using the home source X8PROTEUM (Bruker) with the flexible kappa goniostat. These data allowed tracing two conformations of the C-terminal part up to the residue 239 (C1) and 236 (C2) (Figure 1B). The 20-fold redundancy was necessary to obtain a traceable electron density map for the above residues at low sigma cutoffs (Figure 1B). Further improvement of the data quality did not reveal additional details in electron density maps. In case of TaqSSBA228–252 R190A the model of the flexible C-terminal part starting from residue L253 was built by the following iterative procedure. The model including residues 1–227 was refined by ARP/wARP (31) from CCP4i programme package (32) using 18-fold redundant synchrotron data set (Table 1). The ARP/wARP model phases and figures of merit were used in the following density modification procedure by DM programme (32). During the procedure of density modification the region occupied by the existing protein model and additional region covering all possible directions for the next two C-terminal residues were protected by the mask prepared by MAMA and MAPMAN programmes (33,34). The electron density map obtained after the density modification was used to trace one or two additional residues. The new residues were included into the protein model and the same steps were repeated for the next two residues. Such a procedure allowed building the C-terminal residues up to E260 after which the maps have become untraceable.

---

**Table 1.** Crystallographic data and refinement statistics for wild-type TaqSSB and the TaqSSBA228–252 R190A mutant

<table>
<thead>
<tr>
<th>Protein</th>
<th>TaqSSB wild-type</th>
<th>TaqSSBA228–252 R190A</th>
</tr>
</thead>
<tbody>
<tr>
<td>PDB code</td>
<td>2ihe</td>
<td>2ihf</td>
</tr>
<tr>
<td>Crystal parameters</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Space group</td>
<td>C2</td>
<td>C222</td>
</tr>
<tr>
<td>Cell parameters; a, b, c, α, β, γ</td>
<td>69.8, 80.7, 56.5, 90, 124.4, 90</td>
<td>67.6, 80.4, 97.8, 90, 90, 90</td>
</tr>
<tr>
<td>Data collection</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Beamline/source</td>
<td>ID13, ESRF</td>
<td>BW6, DESY</td>
</tr>
<tr>
<td>Wave length (Å)</td>
<td>0.976</td>
<td>1.07</td>
</tr>
<tr>
<td>Resolution of data (Å)</td>
<td>20.0–2.1</td>
<td>20.0–1.9</td>
</tr>
<tr>
<td>No. of observations/unique reflections</td>
<td>74 521/14 759</td>
<td>391 536/21 312</td>
</tr>
<tr>
<td>Completeness (total/high)%</td>
<td>96.8/95.6</td>
<td>99.6/98.9</td>
</tr>
<tr>
<td>Redundancy</td>
<td>5.1</td>
<td>18.4</td>
</tr>
<tr>
<td>&lt;l/σ(l)&gt;/total/high%</td>
<td>12.2/5.1</td>
<td>20.8/4.4</td>
</tr>
<tr>
<td>R_{sym} (total/high)%</td>
<td>8.0/33.4</td>
<td>10.3/36.7</td>
</tr>
<tr>
<td>Refinement statistics</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Resolution range (Å)</td>
<td>20.0–2.1</td>
<td>20.0–1.9</td>
</tr>
<tr>
<td>No. of protein atoms</td>
<td>1588</td>
<td>1650</td>
</tr>
<tr>
<td>No. of waters</td>
<td>124</td>
<td>185</td>
</tr>
<tr>
<td>R_{work}/R_{free} %</td>
<td>20.9/25.3</td>
<td>22.2/26.1</td>
</tr>
<tr>
<td>r.m.s. deviation for bonds (Å)/angles (deg)</td>
<td>0.007/1.3</td>
<td>0.007/1.4</td>
</tr>
</tbody>
</table>

*For the R_{free} calculations 5% of the total number of reflections were used.

---

TaqSSBA228–252 R190A mutant (3.6 mg/mL in 10 mM Tris–HCl pH 8.0) has crystallized in similar conditions with reservoir containing 100 mM imidazol pH 6.2, 35% (v/v) 2-ethoxyethanol and 220 mM calcium acetate. Triangle-shaped crystal plates of wild-type TaqSSB (0.3×0.3×0.05 mm³) appeared within 2–3 days. Prior to flash cooling, crystals were rinsed in the corresponding reservoir solution.
RESULTS AND DISCUSSION

Structure of wild-type TaqSSB

The 3D structure of the single-stranded DNA-binding protein from *T. aquaticus* (TaqSSB) was determined at 2.3 Å resolution. Figure 2 shows an overall view of the structure. From 264 amino acid residues of the TaqSSB sequence this structure contains residues 2–239 excluding residues 112–126, 162–169 and 205–213 where no electron density was observed. The main part of the protein comprises an elongated β-barrel domain with two adjacent α-helices, which is formed by two OB-folds (9) covalently linked by a flexible loop. The flexibility of the loop is reflected in the lack of electron density for OB-fold connector residues 112–126. The hydrophobic core of the molecule contains 39 residues. Most of these residues (29) are highly conserved in OB-folds of bacterial SSB proteins, which correlates with the high degree of structural conservation in the molecular architecture of SSB proteins. Interestingly residue H54 is part of the conserved hydrophobic core amino acids. It has been shown that the corresponding residue (H55) of the SSB protein from *E. coli* (EcoSSB) is crucial for the formation of the biological active tetramer (35). Mutating H55 of EcoSSB to Y or K destabilizes the OB–OB interaction to such an extent that only monomeric species are observed in solution (36). A number of hydrophobic residues are partially exposed and most of them form a vast hydrophobic patch on one side of the β-barrel in the region where two monomers associate to form the biological dimer (see below). Another source of molecular structure stabilization is provided by a vast network of main-chain hydrogen bonds connecting β-strands into the β-barrel and a number of surface hydrogen bonds and salt bridges between the side chains of charged and polar residues. There are 16 side-chain hydrogen bonds in TaqSSB that stabilize the mutual arrangement of the protein’s secondary structure elements. Most of the residues participating in these interactions are not conserved, which may imply that the corresponding hydrogen bonds were acquired to increase the protein stability in the thermophilic organism. This assumption is supported by the fact that some of the non-conserved surface hydrogen bonds tend to cluster into more extended networks (like interactions E102–R81–E83–R100 or R218–Q177–E220). Formation of such clusters is known to contribute to the stabilization of protein structures from thermophilic and extreme thermophilic organisms (37,38).

A close structural analog of TaqSSB is SSB from *D. radiodurans* (DraSSB) (14) (the r.m.s.d. value for the Cα atoms of the two structures is 1.3 Å). The main differences are observed in the region of the N-terminal α-helix, and in the loop regions. The part of loop 3, which is missing in the DraSSB structure, was traceable in TaqSSB although the electron density is weak. On the other hand the OB-fold connecting region structured as β-hairpin in DraSSB (residues 109–128 in DraSSB) appears to be very flexible and completely disordered in TaqSSB, as indicated by the complete lack of electron density.

Structure of the TaqSSB C-terminal part and of TaqSSB Δ228–252 R190A

Although the structure of a couple of bacterial SSB proteins has been solved (14–19,39), no structural information was obtained about the glycine-rich region and the ultimate C-terminal part. In the case of wild-type TaqSSB, the electron density maps calculated using high-redundancy x-ray
Figure 2. (A) Amino acid sequence and secondary structure of the SSB protein from *Thermus aquaticus*. The N-terminal (underlined in green) and the C-terminal (underlined in red) OB-folds are connected by a flexible loop and form together an elongated β-barrel domain with two adjacent α-helices. The amino acids are coloured according to their similarity to the corresponding residues in EcoSSB. (B) Ribbon representation of the overall structure of wild-type TaqSSB. The colour changes from blue to red from N- to C-terminus, respectively. The two conformations of the glycine-rich C-terminal region are labelled with C1 and C2.

Figure 3. Summary of the structural information on the flexible C-terminal part of TaqSSB obtained from wild-type TaqSSB and TaqSSBΔ228-252 R190A mutant crystal structures. The ball-and-stick model on white background represents residues resulting from the tracing of experimental electron density maps. Residues drawn on light yellow background had no experimental electron density and were obtained by a crystallographic energy minimization procedure. These residues were not included in the final structural model. The Cα-trace of the above mentioned regions is shown in yellow. Orange and magenta ribbons represent the unresolved flexible parts following residues G236 and P239 in the two conformations of the glycine-rich region. A portion of the 0.8σ 2Fo-Fc composite omit map for residues L253–E260 from TaqSSBΔ228-252 R190A obtained as described in ‘Materials and Methods’ section is shown on the right.
diffraction data allowed tracing of two alternative conformations for the glycine-rich region. Chain tracing could be extended to residues 239 (C1) and 236 (C2) at low sigma cutoffs (Figures 1B and 3). The transition between C1 and C2 is coupled with the cis-trans isomerization of proline P231. Due to the high degree of flexibility, no information could be obtained about the structure of the last 25 amino acids. For EcoSSB the last 10 amino acids have been shown to be crucial for the interaction with other proteins involved in DNA replication, repair and recombination (1–5,14) and to be essential for the survival of the bacterial cell (21).

Since this region is separated from the well-structured DNA-binding domain by the flexible glycine-rich region, it was not possible to trace the electron density to the very end of the protein. Therefore, we constructed the deletion mutant TaqSSBΔ228–252, lacking the unstructured glycine-rich region, in which the last 12 amino acids are directly fused to the DNA-binding domain. Unfortunately, this protein formed twinned crystals that were not suitable for x-ray analysis. Mutation of arginine 190 to alanine (TaqSSBΔ228–252 R190A) resolved this problem. Since R190 is involved in intermolecular interactions of wild-type TaqSSB proteins in the crystal, the crystal packing of the TaqSSBΔ228–252 R190A mutant was completely different and no tendency to form twinned crystals could be observed. The TaqSSBΔ228–252 R190A mutant contains 239 amino acids in the sequence. To ease comparison of the structures of wild-type TaqSSB and TaqSSBΔ228–252 R190A the numbering of the residues in the C-terminal part of both proteins is kept identical. Thus, in the sequence of the mutant L253 directly follows P227. The sequence of the mutant thus contains amino acids numbered 1–227 and 253–264. Residues 112–125, 162–168, 205–213 and 261–264 are excluded from the model due to lack of electron density. Except for loop 1, loop 3 and loop 7, the structure of the main part is practically identical for mutant and wild-type TaqSSB (r.m.s.d. for Cα atoms is 0.8 Å). The observed differences in the loop regions are most likely a consequence of the different crystal packings. Due to the lack of the glycine-rich region parts of the C-terminus, which has been elusive to structure determination before, can be observed and are found to be highly disordered and devoid of secondary structure. In the mutant protein structure, electron density extends almost to the C-terminus. Model building was possible only up to the residue E260 at low sigma values (see Figure 3), as electron density becomes very weak and disconnected closer to the C-terminus. Tracing was not possible for the four C-terminal amino acids. The summary of C-terminal structural information obtained from both wild-type TaqSSB and mutant structures is presented in Figure 3. Weak electron density and a considerable amount of noise in those regions reflect a high degree of conformational variability, which implies that the observed structures represent only a small fraction of all possible conformations that can exist in solution. Nevertheless, these results contain important information, since the C-terminal region is responsible for the interaction of SSB with a couple of different proteins such as the χ subunit of DNA polymerase III (4,5), exonuclease I (2) and uracil DNA glycosylase (3). The structural flexibility of this region might be a prerequisite for its functional versatility. The fact that the glycine-rich region is disordered and protrudes from the DNA-binding domain suggests that the last 10 amino acids should be easily accessible for different SSB interaction partners, which differ considerably in size. We tested the protein–protein interaction behaviour of TaqSSB using the χ subunit of E.coli DNA polymerase III as a representative interaction partner. The affinity of this interaction (unpublished data) is similar to the affinity between the χ protein and the cognate EcoSSB (5). By mutational analysis, we showed that for both EcoSSB (5) and TaqSSB (unpublished data) the C-terminus is responsible for this interaction.

Docking of TaqSSB C-terminus to the χ subunit of DNA polymerase III

To examine how such an unstructured region could interact with the χ protein, we docked the peptide E254–F264 in the potential χ–SSB binding site proposed by Gulbis et al. (22). The structure of the χ protein was taken from the crystal structure of the χx subassembly of the E.coli DNA polymerase clamp-loader complex (pdb code: 1em8). The starting peptide conformation was that of the peptide in the TaqSSBΔ228–252 R190A structure. The docking was performed using a combination of flexible peptide docking procedures (40) and molecular dynamics calculations followed by the molecular mechanics energy minimization (27). The result of the docking is presented in Figure 4. The final geometry of the peptide has good stereochemistry and its conformation is unstrained. The highly conserved positively charged residues K124, R128, K132 and R135 of the χ protein interact with the negatively charged and highly conserved residues E259, E260 and D261 from TaqSSB. These interactions lead to the formation of four hydrogen bonds. The side chain of TaqSSB E259 is accommodated in the groove on the surface between K132 and R135 of χ. Residues F256, P257 and L262 of TaqSSB make hydrophobic interactions with residues L141, T143, V117 and L8 of χ. The terminal F264 of TaqSSB makes a stacking interaction with Y119 of χ. This stacking interaction is facilitated by the kink in the amino acid chain, which is caused by the presence of the preceding proline residue. These ultimate two amino acid residues are extremely well conserved in bacterial SSB proteins. In the case of EcoSSB, the mutant strain E.coli ssb113 has been described, in which the penultimate proline residue is replaced by a serine. The importance of this residue is emphasized by the fact that the mutation is conditionally lethal. Cells producing the mutant protein are rendered hyper-sensitive to UV radiation and high temperatures (41,42). In our model the penultimate proline residue is not directly involved in the protein–protein interaction, but is essential for bringing the C-terminal phenylalanine in the correct conformation to form a stacking interaction with Y119 of the χ subunit of E.coli DNA polymerase III.

Biological assembly

The structure of bacterial SSB proteins is an arrangement of four OB-folds. In the TaqSSB monomer two of these folds are covalently linked and two monomers form an OB2–OB2 dimer with 2-fold symmetry. In homotrimeric SSB proteins, like EcoSSB and MtuSSB (SSB from M.tuberculosis), two OB-folds form a dimer and two of these dimers are arranged in a 2-fold symmetry to form the tetramer with
a structure similar to that of the TaqSSB dimer. In TaqSSB
the extensive contact area covering 1397 Å² surface of each
monomer has a flat topology and comprises predominantly
hydrophobic and some polar groups. Comparison of the
mutual arrangement of TaqSSB monomers with arrangements
of the corresponding homodimers of EcoSSB and MtuSSB
reveals that in TaqSSB the subunits are rotated with respect
to each other more than in MtuSSB but less than in EcoSSB
(Figure 5A). Such a rotational variability in the subunit
arrangement of SSBs from different species implies that the
precise orientation of the two subunits may not be crucial for
the protein function. This is supported by modelling of the
TaqSSB DNA complex (see below), which shows that the
difference in orientation of TaqSSB OB₂ subunits compared
to EcoSSB does not produce any hindrance for DNA binding.
Moreover, even larger changes of the OB₂–OB₂ rotation
angle do not create sterical problems for DNA binding.
Nevertheless, revealing the factors defining the OB₂–OB₂
rotation angle will help understanding the diversity in qua-
ternary structures of SSB proteins from different organisms.

Analysis of the OB₂–OB₂ interface of TaqSSB, EcoSSB
and MtuSSB shows that the total buried hydrophobic areas
are 1664 Å², 735 Å² and 2259 Å², respectively. Using
a rough estimate of the free energy of the hydrophobic effect,
which is about +0.105 kJ/mol for each Å² of accessible
hydrophobic surface transferred from non-polar solvent to
water (43), the gain in free energy resulting from the forma-
tion of SSB OB₄ assemblies would be 174 kJ/mol for
TaqSSB, 77 kJ/mol for EcoSSB and 237 kJ/mol for MtuSSB.
In the interface, the hydrophobic residues are oriented in such
a way that the hydrophobic patches of interacting molecules
form an extensive intermolecular hydrophobic core. This
core together with intramolecular cores of the interacting
molecules creates the joint, extended hydrophobic core of
the biologically active OB₄ unit, which is packed so tightly
that it appears like a single molecule. This interaction can
be additionally stabilized by electrostatic interactions or
hydrogen bonds (like hydrogen bonds A2 (main chain)–D105, Q7–Q7, Q130–Q130 in TaqSSB or salt bridge
K7–E80 in EcoSSB). Formation of joint, extended hydropho-
ic cores was previously detected in structures of ribosomal
proteins and their complexes (38,44–46) and it was found
that such a type of interaction is common for some proteins
within the ribosome. To check whether the efficiency of the
joint, extended hydrophobic core formation correlates with the
OB₂–OB₂ rotation angle, we calculated the degree of hydro-
phobic interaction as a function of that angle. The ‘native’
OB₂–OB₂ angles in structures from TaqSSB, EcoSSB and
MtuSSB were considered to be zero. Then for each protein
one of the OB₂ subunits was rotated with respect to the other
around the pseudo-2-fold axis. The rotation range was varied
from 0° to 360° in 2° steps and the solvent-accessible hydro-
phobic surface was calculated for each step. The resulting
angular dependencies are presented in Figure 5B. For both
TaqSSB and MtuSSB, the native OB₂–OB₂ orientations cor-
respond to global minima of the accessible hydrophobic
surface. The minima have sharp parabolic profiles and a
deviation of 10° leads to a noticeable increase in accessible
hydrophobic surface. In case of EcoSSB the curve profile in
the vicinity of the native orientation is not parabolic but
rather box shaped with a broad minimum. If we assume
that the efficiency of the joint, extended hydrophobic core
formation would determine the native OB₂–OB₂ orientation
then the curve profile would mean that EcoSSB possesses a
degree of rotational flexibility around the pseudo-2-fold axis to allow at least 10° rotations. Such rotational flexibility
is observed in the crystal structures of an EcoSSB–ssDNA
complex and of the free protein, where the OB₂–OB₂ angles
differ by 10° (16,47) (Figure 5A). Rotational flexibility in
the case of EcoSSB is also supported by the above free energy
considerations: estimates of the free energy of the OB₂ unit
formation show 2–3 times less gain for EcoSSB com-
pared to TaqSSB or MtuSSB. The same solvent-accessible
hydrophobic surface calculations for an EcoSSB–ssDNA complex (with DNA omitted) show that the native orientation is located in a different region of the broadened minimum. The changes in the broadened minimum and in the absolute values of the solvent-accessible hydrophobic surface are due to reorientation of the hydrophobic surface residues interacting with DNA bases. Fine tuning of the orientation of the EcoSSB OB units within the broadened minimum of orientation angles is then accomplished by additional amino acid interactions and by interactions with ssDNA. The corresponding profiles for TaqSSB and MtuSSB suggest that both these proteins do not change their quaternary structure upon DNA binding.

Modelling the TaqSSB–DNA complex

In fluorescence titrations we found that at high salt conditions the binding of ssDNA to TaqSSB and EcoSSB is very similar (data not shown). Thus, a TaqSSB–DNA complex was modelled by docking the ssDNA structure from an EcoSSB–ssDNA complex (47) (pdb code: 1eyg) to TaqSSB followed by energy minimization of the resulting structure. There are 48 amino acid residues interacting with ssDNA in this model, 13 of them are conserved between TaqSSB and EcoSSB. In a detailed analysis of the amino acids involved in binding ssDNA in this model we did not observe differences in the interaction of the N- and C-terminal OB-fold that could lead to a weaker binding of the N-terminal fold, as found for the related DraSSB (14). The model indicates that ssDNA binds to TaqSSB in a similar way as to EcoSSB. The interactions include hydrogen bonds, Coulomb interactions, hydrophobic contacts and stacking interactions. The modelling shows that with minor adjustments of TaqSSB side chains most stacking interactions described for EcoSSB (47) are preserved. The difference in the mutual orientation of the protein subunits in the biological dimer of TaqSSB and the EcoSSB tetramer does not lead to any hindrance for ssDNA binding. Modelling of the ssDNA onto the TaqSSB structure was not possible with the glycine-rich region being in conformation C1, whereas conformation C2 is compatible with ssDNA binding.

SUMMARY

We have solved the structures of the single-stranded DNA binding protein from the thermophilic organism *T. aquaticus* (TaqSSB) and its deletion mutant TaqSSBΔ228–252 R190A. Both structures showed the main part of the protein to consist of two OB-folds (9) covalently linked by a flexible loop. In the flexible glycine-rich region we observe a bifurcation of the Cα trace at P231 existing as both cis- and trans-isomers. Due to the flexibility of this region, the C-terminus of the wild-type protein could not be resolved. The highly conserved last 10 amino acids play an important role in interactions of SSB proteins with other proteins involved in DNA replication, repair and recombination (1–5,13,21). In the structure of the deletion mutant, in which the glycine-rich region is missing, we could show this highly conserved C-terminal region to be highly disordered and devoid of secondary structure. This might be a prerequisite for the interaction with different protein interfaces presented by different interaction partners.

Docking the structural model obtained for the last 11 amino acids of TaqSSB to the putative SSB binding site of
the χ subunit of DNA polymerase III from *E. coli* (22) showed that the last nine amino acids are able to make a tight interaction with χ. Besides the formation of four hydrogen bonds between highly conserved negatively charged residues of TaqSSB and positively charged amino acids of χ, a stacking interaction between the ultimate phenylalanine residue of TaqSSB and a tyrosine residue of χ can be formed. This interaction is facilitated by the penultimate amino acid of TaqSSB, a proline, which creates a kink in the peptide chain to position the phenylalanine for stacking with the tyrosine of χ. In EcoSSB a mutation of this proline to serine renders the *E. coli* cells UV- and temperature-sensitive (42) and reduces the affinity of EcoSSB to other proteins (2,4).

All bacterial SSB proteins function as a combination of four OB-folds, either as a molecular tetramer with a single OB-fold per molecule or, in the case of the Thermus/Deinococcus group, as a molecular dimer. Whereas the overall structure of the SSB proteins is similar, independent of their oligomeric composition, there exists a significant difference in the angles formed by the two OB3 units interacting to form the biologically active proteins. Hydrophobic interaction calculations show that the formation of an extended hydrophobic core determines the relative orientation of the OB3 units. For EcoSSB our calculations predicted a rotational flexibility which in turn is reflected by the differences between free and ssDNA bound structures (16,47).

**ACKNOWLEDGEMENTS**

We thank Lidia Litz for excellent technical assistance, Dr Joachim Greipel for valuable discussions and Dr Anita Coetzee (Bruker, Delft) for help with the X8PROTEUM high-redundancy data collection and valuable discussions. This work was supported by a HIL grant of Hannover Medical School to U. Curth and by grant MA1081/5-3 from Deutsche Forschungsgemeinschaft. Funding to pay the Open Access publication charges for this article was provided by Hannover Medical School.

**Conflict of interest statement.** None declared.

**REFERENCES**


