Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Structure of the single-stranded DNA-binding protein SSB from *Thermus aquaticus*

Robert Jędrzejczak,[a] Mirosława Dauter,[b] Zbigniew Dauter,[a]* Marcin Olszewski,[c] Anna Długołęcka[c] and Józef Kur[c]*

[a]Synchrotron Radiation Research Section, MCL, National Cancer Institute, Argonne National Laboratory, Argonne, IL 60439, USA, [b]Basic Research Program, SAIC-Frederick, Argonne National Laboratory, Argonne, IL 60439, USA, and [c]Department of Microbiology, Chemical Faculty, Gdańsk University of Technology, Narutowicza 11/12, 80-952 Gdańsk, Poland

Correspondence e-mail: dauter@anl.gov, kur@chem.pg.gda.pl

Received 19 May 2006
Accepted 5 September 2006

PDB Reference: SSB, 2fxq, r2fxqsf.

The crystal structure of the single-stranded DNA-binding protein from *Thermus aquaticus* has been solved and refined at 1.85 Å resolution. Two monomers, each encompassing two oligonucleotide/oligosaccharide-binding (OB) domains and a number of flexible β-hairpin loops, form an oligomer of approximate D_2 symmetry typical of bacterial SSBs. Comparison with other SSB structures confirms considerable variability in the mode of oligomerization and aggregation of SSB oligomers.

1. Introduction

Single-stranded DNA-binding proteins (SSBs) occur in all organisms, from bacteria to eukaryotes. They are involved in the replication, recombination and repair of DNA, preventing it from degradation and the formation of aberrant secondary structures (Greipel *et al.*, 1989). SSB molecules contain a structurally conserved oligonucleotide/oligosaccharide-binding (OB) domain (Murzin, 1993; Suck, 1997). Most bacterial SSBs exist as homotetramers built up of subunits of about 120 amino-acid residues containing the N-terminal OB domain and an unstructured C-terminal tail of about 50–60 residues, which is rich in glycines and prolines. The eukaryotic mitochondrial SSBs are closely analogous. Somewhat different from this characteristic arrangement are the SSBs from thermophilic *Deinococcus* and *Thermus* genera (Dąbrowski, Olszewski, Piątek, Brillowska-Dąbrowska *et al.*, 2002; Dąbrowski, Olszewski, Piątek, Kur *et al.*, 2002; Eggington *et al.*, 2004). These SSBs consist of two consecutive OB domains joined by a short linker and also possess a typical C-terminal tail. They form homodimers containing four OB domains in analogy to the homotetrameric SSBs. In this paper, all these structurally analogous SSBs will be called 4-OB proteins, since all of them contain four OB domains in their stable oligomeric form.

To date, the three-dimensional X-ray structures of 4-OB SSBs from the following sources have been determined: *Escherichia coli* [PDB codes 1kaw (Raghunathan *et al.*, 1997) and 1sru (Savvides *et al.*, 2004)], *E. coli* with bound DNA (PDB code 1eyg; Raghunathan *et al.*, 2000), human mitochondria (PDB code 3ull; Yang *et al.*, 1997), *Mycobacterium tuberculosis* (PDB code 1ue1; Saikrishnan *et al.*, 2003), *M. smegmatis* (PDB code 1x3e, Saikrishnan *et al.*, 2005), *Deinococcus radiodurans* (PDB code 1se8; Bernstein *et al.*, 2004) and *Thermus thermophilus* (PDB code 2cwa; unpublished work).

We have identified, expressed and purified the SSB proteins from *T. thermophilus* and *T. aquaticus* (Dąbrowski, Olszewski, Piątek, Brillowska-Dąbrowska *et al.*, 2002; Dąbrowski,

Olszewski, Piątek, Kur *et al.*, 2002). Here, we describe the X-ray structure of the *T. aquaticus* SSB (TaqSSB) and compare it with those of other 4-OB SSBs.

2. Materials and methods

2.1. Purification, crystallization and data collection

E. coli BL21 cells were transformed with pET-30LIC carrying the SSB-coding sequence from *T. aquaticus*. Transformants were grown at 310 K in Luria–Bertani medium (Miller modification) to an OD_{600} of 0.3 and were induced with IPTG to 0.5 m*M* concentration. The culture was harvested after 12 h and suspended in buffer *A* (50 m*M* Tris pH 7.8, 50 m*M* NaCl). Cells were disrupted by sonication for 20 min at a temperature below 333 K and were centrifuged at 20 000*g* for 1 h. Supernatant was exposed to heat-shock at 343 K for 30 min followed by centrifugation using the conditions described above. The solution was then applied onto a DEAE cellulose column equilibrated with buffer *A*. A sodium chloride step gradient was used for separation. SSB eluted at 0.3 *M* NaCl and this fraction was precipitated with 0.24 g ml^{-1} ammonium sulfate. The pellet was suspended in buffer *A* with 0.3 *M* NaCl and the solution was applied onto a Superdex 200 16/60 GE Healthcare column. Fractions containing almost homogenous SSB were pooled and concentrated on Centricon 10 from Amicon to 15 mg ml^{-1}.

The protein was crystallized using Hampton Research Crystal Screen 1. The best crystals grew in hanging drops when 25 mg ml^{-1} protein solution in 1 *M* NaCl and 10 m*M* Tris–HCl buffer pH 7.9 was mixed in a 1:1 ratio with well solution containing 25–30% MPD, 0.1 *M* acetate buffer pH 4.6 and 20 m*M* $CaCl_2$. They belong to space group $C222_1$, with unit-cell parameters a = 51.14, b = 163.77, c = 60.16 Å.

1.85 Å resolution diffraction data were collected at SER-CAT beamline BM22 (Advanced Photon Source, Argonne) using a MAR 225 CCD detector from a crystal of dimensions 0.2 × 0.2 × 0.25 mm cryocooled at 100 K after a short soak in mother liquor containing an additional 25% glycerol. 300 diffraction images of 1° width were processed and intensities merged with *HKL*-2000 (Otwinowski & Minor, 1997). The resulting diffraction data statistics are presented in Table 1.

2.2. Molecular replacement and refinement

Most of the programs used were from the *CCP*4 suite (Collaborative Computational Project, Number 4, 1994). The structure was solved by molecular replacement with *AMoRe* (Navaza, 1994). The subunit (both OB domains) of *D. radiodurans* SSB (Bernstein *et al.*, 2004; PDB code 1se8) was used as a search model, pruned of the loops extending from the core of the protein. To minimize the model bias, the phases obtained from the initial MR solution were input to *ARP*/*wARP* (Perrakis *et al.*, 1999), which

Table 1
Statistics of diffraction data and structure refinement.

Values in parentheses are for the highest resolution shell.

Data collection	
Space group	$C222_1$
Unit-cell parameters	
a (Å)	51.14
b (Å)	163.77
c (Å)	60.16
Wavelength (Å)	1.039
Resolution limits (Å)	40–1.85 (1.92–1.85)
Measured reflections	133398 (14625)
Unique reflections	22066 (2165)
Multiplicity	6.0 (6.0)
Completeness (%)	100.0 (100.0)
R_{merge} (%)	5.0 (83.2)
$I/\sigma(I)$	33.9 (2.4)
Refinement	
Resolution (Å)	30.0–1.85
R factor (%)	19.22
All reflections	20912
R_{free} (%)	24.73
Free reflections	1130
R.m.s.d. bond lengths (Å)	0.021
R.m.s.d. bond angles (°)	1.96
PDB code	2fxq

Figure 1
Sequence alignment of SSBs based on their structures. Taq, *T. aquaticus* (this work); Tth, *T. thermophilus* (PDB code 2cwa); Drad, *D. radiodurans* (PDB code 1se8); Dcoli, *E. coli* complexed with DNA (PDB code 1eyg); Mtub, *M. tuberculosis* (PDB code 1ue1); Msmeg, *M. smegmatis* (PDB code 1x3e); Hmt, human mitochondrial (PDB code 3ull). N and C represent the N- and C-terminal domains. The structures were overlapped on 48 C^{α} atoms from each SSB domain (marked in red) which are closer together than 2.0 Å in all models after superposition. Residues not visible in the electron-density maps are in lower case letters.

automatically built 120 amino-acid residues in several chains. This model was then refined with *REFMAC*5 (Murshudov *et al.*, 1997) in a number of iterations interspersed with further manual model building with the use of *QUANTA* (Accelrys, San Diego, USA). When the *R* factor and $R_{\rm free}$ reached 26.4 and 32.2%, respectively, the TLS option in *REFMAC*5 (Winn *et al.*, 2001) was switched on, with the rigid-body fragment encompassing the whole protein core but excluding fragments of the extended loops. This procedure significantly improved the agreement between the observed and calculated amplitudes and the final value of the *R* factor was 19.2% and that of $R_{\rm free}$ was 24.7%. The refinement statistics are summarized in Table 1.

2.3. Superposition of SSB models

All superpositions of various SSB models were performed on C^{α} atoms using *LSQKAB*, selecting analogous stretches of the main chain within the core of the individual OB domains in each of the SSB molecules. The following SSB models were used: *T. aquaticus* from the current work (Taq), *T. thermophilus* (Tth; PDB code 2cwa), *D. radiodurans* (Drad; PDB code 1se8), *E. coli* from the complex with DNA (Dcoli; PDB code 1eyg), *M. tuberculosis* (Mtub; PDB code 1ue1), human mitochondria (Hmt; PDB code 3ull) and *Escherichia coli* without DNA (Ecoli; PDB code 1kaw).

Taking into account that each Taq and Drad subunit consists of two OB domains and that there are four independent monomers in the Dcoli structure, two independent monomers in Mtub and in Hmt and four identical monomers in the Ecoli structure, there are a total of 13 unique models of individual OB domains available. They were all superimposed onto the N-terminal domain of Taq (TaqN) using those pairs of C^{α} atoms that do not differ by more than 2 Å in each of the 13 overlapped domains. These 48 structurally corresponding residues are shown in red in Fig. 1.

The solvent-accessible area buried by the formation of the oligomers was calculated with *AREAIMOL* by subtracting the area obtained for the whole molecule from the sum of the values obtained for two individual subunits.

3. Results and discussion

The crystal structure of TaqSSB was solved by molecular replacement using the core of the molecule of DradSSB as a search model. The structure of TaqSSB is similar to that of DradSSB and to those of other 4-OB SSBs. In brief, there is one TaqSSB subunit in the asymmetric unit of the crystal and it consists of the core spanning two OB domains, each having the typical five-stranded β-barrel structure decorated with a three-turn α-helix on one side. The N-terminal domain extends from residues 1 to 108 and the C-terminal domain from 129 to 229. Both OB domains are joined by a β-hairpin linker (residues 109–128) and effectively form a larger ten-stranded β-barrel, in analogy to all other 4-OB SSBs. Characteristically for SSBs, several structural features (β-hairpin loops and a C-terminal tail) extend away from the barrel core of the protein and similarly to the case in the other crystals of SSBs some of them are not visible in the electron density. The missing residues in TaqSSB are two at the N-terminus (residues 1–2), nine in loop 2 of the N-terminal OB domain (residues 86–94), 14 in the linker between the two OB domains (residues 111–124), seven in loop 1 of the C-terminal OB domain (residues 163–169) and the long C-terminus from residues 230 to 301. The complete sequence of TaqSSB was present in the crystallized protein. The extended loops 1 and 3 of the N-terminal domain and loops 2 and 3 of the C-terminal domain have well defined conformation, stabilized by the intermolecular crystal contacts. The slightly convex surface created by six antiparallel strands of the large β-barrel faces the analogous surface of the other, twofold symmetry-related subunit, so that those two subunits form an arrangement of four OB domains typical of all bacterial SSBs.

All main-chain torsion angles in the TaqSSB model lie in the allowed regions of the Ramachandran plot. For several amino acids in the vicinity of the missing loops only the main

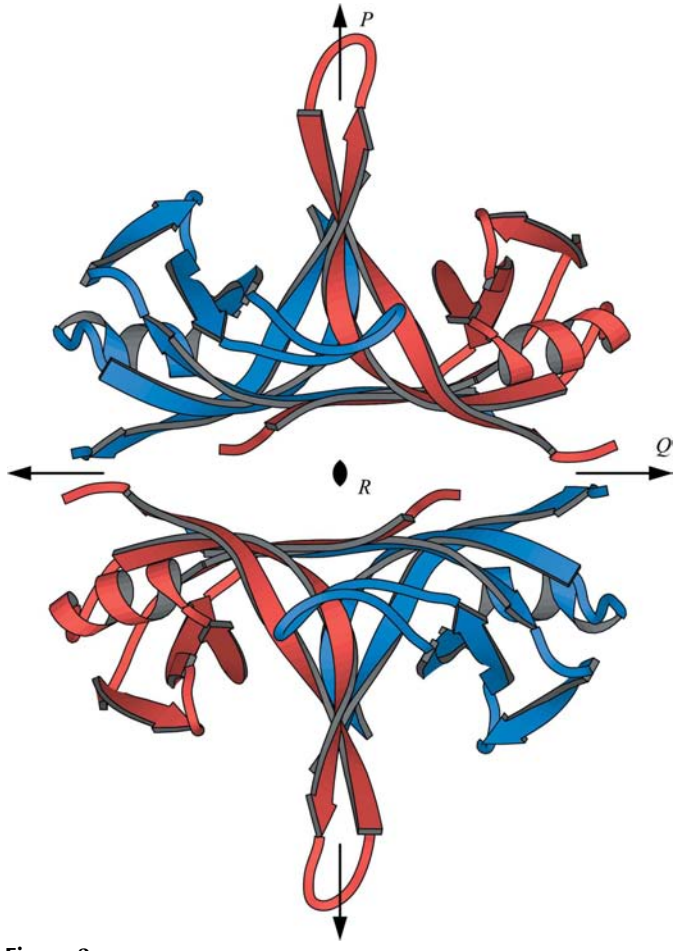

Figure 2
Molecule of TaqSSB viewed along the crystal *y* axis. In the two monomers participating in the formation of the dimeric molecule, the N-terminal domains are shown in blue and the C-terminal domains in red. The molecule has approximate 222 symmetry and its local dyads are marked as *P*, *Q* and *R* according to Saikrishnan *et al.* (2003). The *R* axis coincides with the crystal twofold symmetry axis along the *y* direction; the *P* and *Q* axes are approximate and relate the N-terminal domains to the C-terminal domains within the same or a neighbouring subunit.

Table 2
Architecture of SSB structures.

All molecules were positioned according to their own approximate or exact 222 symmetry and their β-sheets were projected on the plane perpendicular to the P axis. The first column gives the angle φ between the average directions of the core β-sheets in the two participating subunits, as illustrated in Fig. 2. The second column gives the total solvent-accessible area shielded by formation of the 4-OB molecule from two monomers (for Taq, Tth and Drad) or dimers (for all other SSBs) across the Q or R axes.

	φ (°)	Area (Å^2)
Taq	63.6	2523
Tth	63.2	3287
Drad	64.0	2150
Hmt	42.3	2295
Dcoli	36.4	1914
Ecoli1	38.0	1826
Ecoli2	36.1	1658
Mtub	88.4	5786
Msmeg	88.1	5683

chain was built, since their side chains lacked electron density. These are all long charged side chains, mostly arginines (residues Arg3, Arg85, Arg110, Arg125, Arg126, Arg162, Arg170 and Glu161). In spite of the fact that a substantial part of the molecule, 67 of 264 residues (about 25%) are missing in the refined model, the final $R/R_{\rm free}$ factors of 19.2/24.7% are not very high. This confirms that the missing atoms are disordered and do not contribute significantly to the diffraction intensities.

3.1. Symmetry of the oligomer

The single subunit of TaqSSB contains two OB domains arranged in an approximately symmetric fashion. Of the 48 residues forming the most structurally conserved core of each OB domain (see §2.3 above), 27 are different (see Fig. 1), but despite these differences in the sequence the core parts of the two OB domains are highly similar. They are related by the local twofold axis perpendicular to the surface of the extended β-sheet and can be superimposed with an r.m.s.d. of 0.58 Å for 48 C$^\alpha$ atoms of the core. This axis corresponds to the P axis of the oligomer, according to the nomenclature of Saikrishnan *et al.* (2003). This is analogous to the arrangement of the two OB domains in DradSSB, TthSSB and all other known 4-OB SSB structures. The 48 C$^\alpha$ atom pairs of the superposed core β-sheets of all SSBs do not differ by more than 2 Å, regardless of whether the two participating OB domains are in the same or different chains. The bottom surface of the sheet is convex, as seen viewed along the R axis in Fig. 1, and its curvature is similar in all SSB structures, as evidenced by the effective overlap of its core fragments. Not surprisingly, parts of the protein which extend from the β-barrel do not obey the internal twofold symmetry so well.

In the TaqSSB crystal structure two subunits lie on two sides of the crystallographic twofold axis (oligomeric R axis), with their extended antiparallel β-sheets facing each other (Fig. 2). This is analogous to the domain arrangement in all 4-OB SSBs, but the mutual disposition of the two β-sheets differs considerably in different structures. As seen in Fig. 3, the four strands in the central part of the β-surface are antiparallel with respect to each other in all SSBs. However, the angle between the average directions of these strands (marked in Fig. 3 as black lines) in the two stacked β-sheets differs considerably (Table 2). This angle in Taq is about 63°, similar to that in Tth and Drad and intermediate between the values for Ecoli and Hmt (about 40°) and for Mtub and Msmeg (about 88°).

The interaction between the TaqSSB subunits buries 2523 Å^2 of the total surface area during formation of the oligomer. This is more than in the analogous interaction in Ecoli or Drad, but much less than in Mtub and Msmeg (Table 2). This interaction in mycobacterial SSBs is exceptional, since most of the contribution is from the clamp formed by the intertwined C-terminal strands (residues 112–120) of the two subunits at the extreme Q-axis side of the molecule, a feature not observed in other SSBs. The surface buried in this interaction in Taq is intermediate between that in Tth and Drad; it is 15% larger than in Tth and 30% smaller than in Drad. These differences in the tightness of the packing of the molecules are rather surprising for these three very closely related proteins. However, whereas the central parts of the β-sheet in all three (Taq, Tth and Drad) SSBs are very similar, the fragments extending towards the molecular Q axis have somewhat different conformations and these changes are mainly responsible for the observed differences in the extent of the interaction between subunits in these three oligomers. These variable fragments in Taq are at the beginning and end of each OB domain, residues 3–5, 107–110, 125–128 and 227–229, located immediately before or after the disordered N-terminus, linker and C-terminus.

Among the residues important for the interaction of SSB with the 35-mer of single-stranded DNA in the only known structure of a complex, Ecoli (PDB code 1eyg; Raghunathan *et al.*, 2000), are Trp40, Trp54 and Phe60 in one monomer and Trp54 and Phe60 in the second monomer within the dimer wrapped up by the DNA chain. Their aromatic rings form stacking interactions with the DNA bases. The analogous residues in the sequence of Taq are Phe40, Tyr53 and Leu59 in the N-terminal domain and Arg162, Phe173 and Trp179 in the C-terminal domain. Four out of these five residues of Ecoli are also aromatic in Taq and the lack of an aromatic amino acid in position 162 agrees with the asymmetry in the interactions with DNA of the two domains in Ecoli. It is remarkable that two tryptophans, Trp87 and Trp206, are highly conserved in all domains of Taq, Tth, Drad and Ecoli and are either Tyr or Phe in Mtub, Msmeg and Hmt. In the Ecoli complex these tryptophans reside in the close vicinity of DNA in both domains, but are not modeled in the parallel stacking interactions with the bases.

Several basic residues in the vicinity of DNA in the Ecoli complex are conserved as Arg or Lys in at least one domain of Taq, as well as in Tth, Drad and other SSBs. They are arginines 21, 48, 55, 81, 85, 180, 190, 200, 213, 214 and lysines 93, 94, 191. Most probably, these residues stabilize DNA by ionic or hydrogen-bond interactions. It is, however, not certain which other residues of Taq may interact with the DNA, since the

extended loops in Taq differ in conformation from Ecoli or are in part completely disordered.

3.2. Crystal packing

The packing of molecules in the crystal of TaqSSB is illustrated in Fig. 4(*a*). Each molecule is in contact with 12 neighbors, four related by full unit-cell translations along the *x* and *z* directions and eight related by cell *C*-centering translations. The most extensive interactions are formed across twofold axes in two regions. The first such region is formed by residues in loop 3 of the C-terminal domain, where two equivalent β-hairpins form a four-stranded antiparallel β-sheet encompassing residues 205–209; in addition, there is a salt bridge between Glu212 of loop 3 and Arg154. The other region is at the edge of the molecule, where Glu71 from the N-terminal domain forms a salt bridge with Arg226 positioned close to the end of the C-terminal domain. The long β-hairpin loop 3 of the C-terminal domain of TaqSSB is visible in the electron density in its entirety, in contrast to the analogous regions in the N-terminal domain of TaqSSB and in DradSSB, where there are no intermolecular interactions involving residues from this loop.

The packing of tetramers and interactions between them in TaqSSB are different to those in DradSSB, where the intermolecular contacts are formed by other regions, as depicted in Fig. 4(*b*). In DradSSB the most extensive interactions exist across the twofold axis between residues within and in the

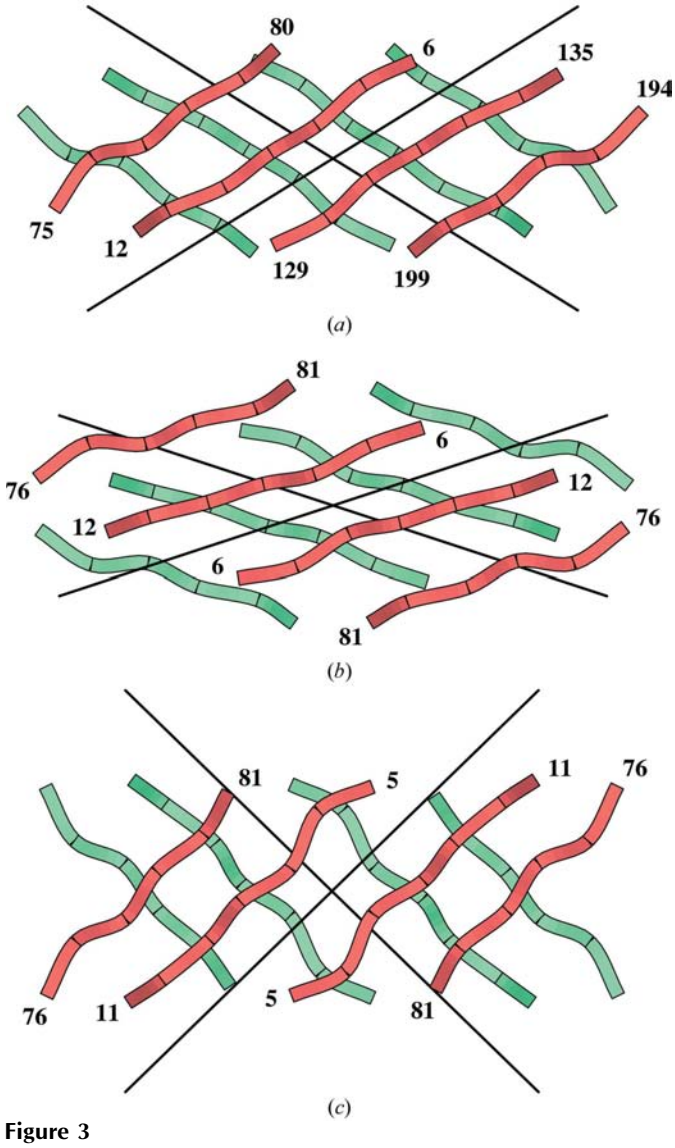

Figure 3
The antiparallel fragments of the core β-strands (residues 6–12 and 70–75 in the TaqN domain and structural equivalents in other SSBs domains) projected along the *P* axis of the molecule for Taq (top), Dcoli (middle) and Mtub (bottom). The black lines show average directions of the strands in the two participating β-sheets. The angles φ between them are presented in Table 2.

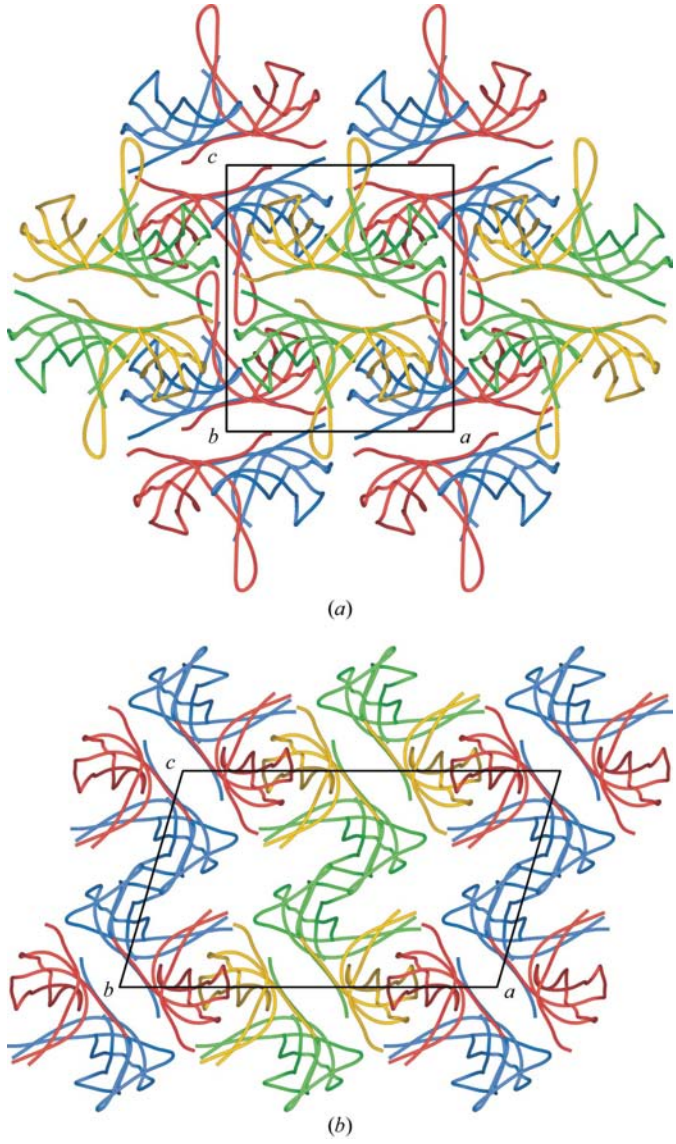

Figure 4
Packing of molecules in the crystal structures of Taq (*a*) and Drad (*b*) SSBs. The N-terminal domains are in blue and green and the C-terminal domains are in red and yellow. The blue/red molecules form one layer in the *xz* plane, related by the twofold axis parallel to the crystal *y* axis (of $C222_1$ symmetry for Taq and of *C*2 symmetry for Drad); the green/yellow molecules form another layer, related by the 2_1 axis with respect to the previous layer.

vicinity of loop 1 of the N-terminal domain, including a salt bridge between Arg21 and Glu31, and across the 2_1 axis between the residues near loop 1 of the C-terminal domain, including a salt bridge between Asp151 and Lys195.

Bernstein *et al.* (2004) postulated that the extensive intermolecular interactions between residues near loop 1 of the N-terminal domain in DradSSB have biological relevance, playing a role in the survival of bacteria from *Deinococcus*/*Thermus* genera under extreme conditions. The crystal structure of TaqSSB, with its completely different arrangement and interactions between molecules, does not support this proposal.

4. Conclusions

The structure of *T. aquaticus* SSB is analogous to all other known structures of bacterial SSBs and is similar to the structures of the *T. thermophilus* and *D. radiodurans* proteins, although there are some significant differences, particularly in the interactions between their individual OB domains and between molecules in the crystal lattice. In the crystal of TaqSSB several loop fragments are disordered, namely those which do not participate in the intermolecular interactions. Similarly to the case in other structures of SSBs, the loops extending from the protein core are well defined in the electron density when they are stabilized by contacts with neighboring molecules in the crystal. It is possible that such flexibility of the loops is important for effective binding of DNA, which wraps around the SSB molecules. Since only one structure of the SSB–DNA complex is known (for Ecoli), confirmation of to what extent the flexible loops adjust to accommodate DNA must wait for the elucidation of more structures of SSB–DNA complexes.

This work was supported in part by the Polish State Committee for Scientific Research (project No. P04A 054 25 to JK) and in part by Federal funds from the National Cancer Institute, National Institutes of Health under contract No. NO1-CO-12400 and by the Intramural Research Program of the NIH, National Cancer Institute, Center for Cancer Research. Use of the SER-CAT beamline at the Advanced Photon Source X-ray facility was supported by the US Department of Energy, Office of Science, Office of Basic Energy Sciences under Contract No. W-31-109-Eng-38. The content of this publication does not necessarily reflect the views or policies of the Department of Health and Human Services, nor does mention of trade names, commercial products or organizations imply endorsement by the US or any other Government.

References

Bernstein, D. A., Eggington, J. M., Killoran, M. P., Cox, M. M. & Keck, J. L. (2004). *Proc. Natl Acad. Sci. USA*, **101**, 8575–8580.
Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.
Dąbrowski, S., Olszewski, M., Piątek, R., Brillowska-Dąbrowska, A., Konopa, G. & Kur, J. (2002). *Microbiology*, **148**, 3307–3315.
Dąbrowski, S., Olszewski, M., Piątek, R. & Kur, J. (2002). *Protein Expr. Purif.* **26**, 131–138.
Eggington, J. M., Haruta, N., Wood, E. A. & Cox, M. M. (2004). *BMC Microbiol.* **4**, 2.
Greipel, J., Urbanke, C. & Maass, G. (1989). *Protein–Nucleic Acid Interaction*, edited by W. Saenger & U. Heinemann, pp. 61–86. London: Macmillan.
Murshudov, G. N., Vagin, A. A. & Dodson, E. J. (1997). *Acta Cryst.* D**53**, 240–255.
Murzin, A. G. (1993). *EMBO J.* **12**, 861–867.
Navaza, J. (1994). *Acta Cryst.* A**50**, 157–163.
Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
Perrakis, A., Morris, R. & Lamzin, V. S. (1999). *Nature Struct. Biol.* **6**, 458–463.
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.
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.
Savvides, S. N., Raghunathan, S., Fütterer, K., Kozlov, A. G., Lohman, T. M. & Waksman, G. (2004). *Protein Sci.* **13**, 1942–1947.
Suck, D. (1997). *Nature Struct. Biol.* **4**, 161–165.
Winn, M. D., Isupov, M. N. & Murshudov, G. N. (2001). *Acta Cryst.* D**57**, 122–133.
Yang, C., Curth, U., Urbanke, C. & Kang, C. (1997). *Nature Struct. Biol.* **4**, 153–157.