Structure of the gene 2.5 protein, a single-stranded DNA binding protein encoded by bacteriophage T7

Thomas Hollis, James M. Stattel, Dane S. Walther, Charles C. Richardson, and Tom Ellenberger*

Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, MA 02115

Contributed by Charles C. Richardson, June 21, 2001

The gene 2.5 protein (gp2.5) of bacteriophage T7 is a single-stranded DNA (ssDNA) binding protein that has essential roles in DNA replication and recombination. In addition to binding DNA, gp2.5 physically interacts with T7 DNA polymerase and T7 primase-helicase during replication to coordinate events at the replication fork. We have determined a 1.9-Å crystal structure of gp2.5 and show that it has a conserved OB-fold (oligosaccharide/oligonucleotide binding fold) that is well adapted for interactions with ssDNA. Superposition of the OB-folds of gp2.5 and other ssDNA binding proteins reveals a conserved patch of aromatic residues that stack against the bases of ssDNA in the other crystal structures, suggesting that gp2.5 binds to ssDNA in a similar manner. An acidic C-terminal extension of the gp2.5 protein, which is required for dimer formation and for interactions with the T7 DNA polymerase and the primase-helicase, appears to be flexible and may act as a switch that modulates the DNA binding affinity of gp2.5.

The chemically reactive bases of DNA are normally sequestered from the intracellular milieu by stacking interactions within the double helix; in higher organisms, protection is also derived from chromatin assembly. Nonetheless, single-stranded DNA (ssDNA) has a transitory existence in cells, arising as an intermediate during DNA replication, repair, recombination, and transcription. Cells have evolved one class of proteins, the ssDNA binding proteins (1), that bind ssDNA in preference to double-stranded DNA (dsDNA). The function of ssDNA binding proteins appears to be 3-fold. They transiently coat and protect ssDNA intermediates in the processes listed above, they eliminate secondary structure in ssDNA and thereby facilitate pairing of homologous strands, and they recruit the proteins of DNA metabolism and coordinate their activities.

Bacteriophage T7, like other viruses that encode their own proteins of DNA metabolism, encodes a ssDNA binding protein that is the product of gene 2.5. Gene 2.5 protein (gp2.5) is functionally similar to the extensively studied SSB protein of *Escherichia coli* and the gene 32 protein of bacteriophage T4 (1, 2). In their respective systems, these proteins are essential for DNA replication and have roles in DNA recombination and repair (1–11). Gene 2.5 protein stimulates DNA polymerase activity and increases the efficiency of RNA primer synthesis through physical interactions with T7 DNA polymerase (9) and T7 primase-helicase (12), respectively. These interactions are undoubtedly involved in the critical role of gp2.5 in coordinating leading and lagging strand synthesis at the replication fork (11). In addition to these roles in replication, gp2.5 has critical roles in DNA recombination and repair (1, 3, 4, 13).

T7 gp2.5, *E. coli* SSB protein, and T4 gene 32 protein contain regions of functional homology, despite a near lack of sequence homology (14). Limited proteolysis of each protein generates several stable fragments suggestive of folded structural domains joined by protease-sensitive segments (1). Each protein has an acidic carboxyl terminus that facilitates interactions with other DNA replication and recombination proteins (1, 6). Aromatic residues within the amino terminus of each protein contribute directly to ssDNA binding (1, 8, 13).

Although gp2.5 is functionally homologous to *E. coli* SSB protein and T4 gene 32 protein, it has a number of distinguishing characteristics. gp2.5 is a stable protein dimer off DNA, with subunit interactions involving the carboxyl-terminal acidic region (6). In contrast, the native form of *E. coli* SSB protein is a tetramer, and T4 gene 32 monomers self-associate into larger aggregates as the protein concentration is increased (1, 8). These differences in subunit interactions are manifested in the different DNA binding behaviors of these ssDNA binding proteins. *E. coli* SSB protein (8, 15) and T4 gene 32 protein (16, 17) bind to DNA cooperatively under low salt conditions, a finding consistent with the formation of a protein filament that is entwined with the ssDNA (18). In contrast, T7 gp2.5 binds to DNA with little cooperativity at any salt concentration (2), suggestive of DNA binding by monomers or stable dimers without additional interactions between subunits on DNA. Under low salt conditions, gp2.5 binds to ssDNA with only one-tenth of the affinity exhibited by the other two proteins (2). When compared with other prokaryotic proteins known to be involved in recombination, such as *E. coli* RecA protein, *E. coli* SSB protein, and T4 gene 32 protein, gp2.5 is considerably more efficient at mediating homologous base pairing (refs. 2 and 3; S. Tabor and C.C.R., unpublished observations). This strand annealing activity (3), coupled with the transfer of homologous strands by the T7 helicase (5), contributes to genetic recombination during growth of T7 phage.

Genetic experiments have shown that neither *E. coli* SSB protein nor T4 gene 32 protein can functionally replace gp2.5 protein *in vivo* (13). *In vitro*, T4 gene 32 protein has only a minimal stimulatory effect on the activity of T7 DNA polymerase (9), and T7 primase-helicase cannot load onto ssDNA that is coated with T4 gene 32 protein, a process that is not impeded by gp2.5 bound to DNA (4). In contrast, *E. coli* SSB protein both stimulates T7 DNA polymerase activity (9) and permits T7 primase-helicase to load onto ssDNA (3, 10). However, gp2.5 causes a 10-fold increase in the frequency of initiation by T7 primase-helicase, whereas *E. coli* SSB protein does not (11). The unique DNA binding characteristics of gp2.5 and its interactions with other replication proteins together account for its irreplaceable roles in T7 replication.

The essential protein–protein interactions of gp2.5 are mediated in part by its acidic C-terminal region. Deletion of 21-aa residues from the carboxyl terminus of gp2.5 results in a truncated protein, gp2.5Δ21C, that cannot support phage growth, form dimers, interact with T7 DNA polymerase or primase-helicase (6), or coordinate the synthesis of leading and lagging strands at an *in vitro* replication fork (11). Protein chimeras in which the C-terminal acidic region of either *E. coli* SSB protein or T4 gene 32 protein is fused to the truncated gp2.5Δ21C protein can support the growth of T7 phage lacking gp2.5 (6).

Abbreviations: ssDNA, single-stranded DNA; gp2.5, gene 2.5 protein; OB-fold, oligosaccharide/oligonucleotide binding fold.

Data deposition: The atomic coordinates have been deposited in the Protein Data Bank, www.rcsb.org (PDB ID code 1JE5).

*To whom reprint requests should be addressed. E-mail: tome@hms.harvard.edu.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

BIOCHEMISTRY

Conversely, a truncated *E. coli* SSB protein or T4 gene 32 protein fused to the C-terminal acidic segment of gp2.5 does not support growth of T7 phage, even though these protein chimeras dimerize and can interact with T7 DNA polymerase and the primase-helicase to stimulate their activities *in vitro* (6). The properties of these chimeric ssDNA binding proteins suggest that gp2.5's acidic tail mediates nonspecific protein–protein interactions, and that the remaining N-terminal region of gp2.5 makes specific interactions with other proteins during DNA replication.

It is notable that gp2.5Δ21C binds to ssDNA with higher affinity than wild-type gp2.5 (J.M.S. and C.C.R., unpublished observations). This increase in binding affinity could result from an increased concentration of monomers, which might be the species that binds to ssDNA. Alternatively, the negatively charged C terminus of gp2.5 could weaken DNA binding affinity, either directly by competing with the DNA for the DNA binding surface of gp2.5, or indirectly by unfavorable electrostatic interactions that promote dissociation of the bound DNA. The C-terminal acidic tail of gp2.5 also contributes to interactions with other T7 replication proteins, as discussed above. The dual roles of the C-terminal tail of gp2.5 in modulating DNA binding affinity and mediating protein–protein interactions suggests the possibility of a switching mechanism that regulates the DNA binding strength of gp2.5 in response to binding other proteins of a DNA replication complex.

To better understand the unique features of the gene 2.5 protein that underlie its functions in T7 replication and recombination, we have determined a crystal structure of the gp2.5 dimer with x-ray data extending to 1.9 Å resolution. The core of the gp2.5 structure is a five-stranded β-barrel, termed an OB-fold (oligosaccharide/oligonucleotide binding fold; ref. 19), which is a conserved feature of many ssDNA binding proteins from bacteria, mammalian cells, and viruses (20–23). The structural similarity of gp2.5 and SSB proteins is remarkable considering the lack of sequence homology, and it is suggestive of a strong evolutionary selection for the barrel-shaped OB-fold as an efficient scaffold for sequestering ssDNA. Several unique structural features on the surface of the OB-fold create the interface between subunits of the gp2.5 dimer, and other unique features are likely to contribute to its specific interactions with T7 replication proteins. A conserved patch of residues on the surface of the OB-fold is the probable location of the ssDNA binding surface.

Materials and Methods

Protein Expression and Crystallization. A gene fragment encoding the truncated protein gp2.5Δ26C was cloned into the pET 19b vector (Novagen, Madison, WI) in frame with a DNA sequence encoding an 8-aa rhinovirus 3C protease recognition sequence and a polyhistidine affinity tag. The resulting construct has an N-terminal histidine affinity tag that can be cleaved with PreScission Protease (Amersham Pharmacia) after purification of gp2.5Δ26C. Selenomethionyl gp2.5Δ26C was produced in *E. coli* by published methods (24, 25), and purified by nickel chelate affinity chromatography. The PreScission Protease was added to the purified gp2.5Δ26C protein at a ratio of 1:100 and the mixture was dialyzed overnight at 4°C against Tris·Cl, pH 7.5/300 mM NaCl/3 mM DTT. The protease was then removed by passing the reaction mixture over a 1-ml glutathione agarose column (Amersham Pharmacia).

gp2.5Δ26C was crystallized by the hanging drop vapor diffusion method. Equal volumes of a protein solution (12 mg/ml) and a well solution containing 22% PEG 4000/100 mM Tris·Cl, pH 7.0/250 mM $CaCl_2$/14% ethylene glycol were mixed together and incubated at 22°C. Crystals of gp2.5Δ26C appear after 2 days that belong to space group $P2_12_12_1$ and have two protein molecules in the asymmetric unit. The dimensions of the unit cell are a = 68.1 Å, b = 71.8 Å, and c = 82.2 Å.

Structure Determination. Crystals of selenomethionyl gp2.5Δ26C were soaked with 1 mM phenylmercuric acetate in the crystallization well solution for 3 days. X-ray data from a single crystal were collected by using energies corresponding to the selenium K-edge (λ = 0.9792 Å) and the mercury L_{III}-edge (λ = 1.0073 Å) at beamline X-25 of the National Synchrotron Light Source (Upton, NY). Native x-ray data were collected at beamline A1 of the Cornell High Energy Synchrotron Source (Ithaca, NY) and x-ray intensity data were processed with DENZO/SCALEPACK (26). Six mercury binding sites and four well ordered selenium sites were located in the crystallographic asymmetric unit by anomalous difference Fourier methods. The heavy atom parameters for these metals were refined and single-wavelength anomalous diffraction (SAD) phases were calculated with the program MLPHARE (27). The best electron density map was obtained after density modification by using the program RESOLVE (28). Model building was performed with the program O (29). The model was refined against the native data set by conjugate gradient minimization and torsion angle-restrained molecular dynamics using the program CNS (30). The success of the refinement was evaluated at each stage by the change in the free *R* factor (31) and by inspection of stereochemical parameters with either the PROCHECK (32) or the ERRAT program (33). Model refinement converged with a final *R* factor of 21.4% (R_{free} = 26.8%), using all observed x-ray measurements in the resolution range 25–1.9 Å. The coordinates of the gp2.5 dimer have been deposited in the Protein Data Bank (PDB ID code 1JE5).

Protein Sequence and Structure Alignments. The structural coordinates of *E. coli* SSB protein (1EYG) and human RPA70 (1JMC) bound to DNA were obtained from the Protein Data Bank (http://www.rcsb.org/pdb) and superimposed onto the gp2.5Δ26C model coordinates with program DALI (34, 35; http://www2.ebi.ac.uk/dali/dali.html). A least-squares superposition of the protein models that was based on the structurally homologous residues identified by DALI was then performed with the modeling program O (29).

Results and Discussion

Structure of gp2.5. We crystallized a truncated form of T7 gp2.5 termed gp2.5Δ26C, which lacks 26 residues at its C terminus. Full-length gp2.5 did not crystallize, perhaps because its C-terminal acidic segment is not stably folded and may interfere with crystal packing. gp2.5Δ26C binds to ssDNA with slightly higher affinity than wild-type gp2.5, but it does not stably dimerize or interact with T7 DNA polymerase or the T7 primase-helicase (J.M.S. and C.C.R., unpublished results). Although the dimerization activity of gp2.5Δ26C is impaired, it crystallized with two protomers in the crystallographic asymmetric unit. The structure of gp2.5Δ26C was determined by single wavelength anomalous diffraction experiments at two different x-ray energies (Fig. 1*A*, Table 1), using a crystal of the selenomethionyl protein that was equilibrated in phenylmercury acetate. After density modification by solvent flattening (36), the electron density was of high quality for most residues in the asymmetric unit (Fig. 2). Several disordered loops, spanning residues 25–33 and 79–85 of both protomers and residues 183–188 of one protomer, were omitted from the crystallographic model. The root-mean-square deviation of all backbone atoms of the two molecules in the asymmetric unit is 0.34 Å, with a larger than average deviation of the C-terminal region resulting from two different crystal packing interactions. The crystallographic model of the gp2.5 dimer has been refined to an *R* factor of 0.21 (R_{free} = 0.26), using all x-ray data extending to a resolution limit of 1.9 Å (Table 1).

gp2.5Δ26C is a compact single domain protein consisting of a five-stranded anti-parallel β-barrel, the OB-fold (19), capped by

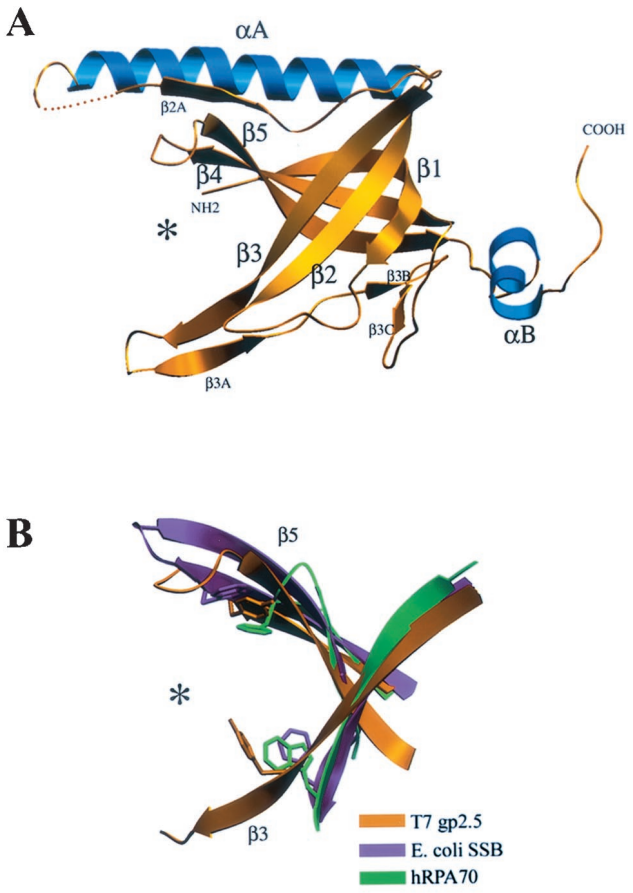

Fig. 1. gp2.5 is an OB-fold protein. (*A*) The structure of gp2.5 is a single domain consisting of a five-stranded anti-parallel β-barrel (the OB-fold; gold) capped by a α-helix (αA; blue). A 28-aa tail protrudes away from the β-barrel at the carboxyl-terminus of the gp2.5Δ26C protein. An additional 26 residues, which are missing from the C terminus of gp2.5Δ26C, participate in dimer formation and interactions with other replication proteins. (*B*) The proposed DNA binding surface (asterisk) contains a pair of structurally conserved aromatic residues surrounded by basic residues. This region of the gp2.5 structure (gold) is superimposed on structures of the *E. coli* SSB protein (magenta) (18) and human RPA70 (green) (37). In the *E. coli* SSB protein and human RPA70 structures these residues contact three nucleotides of the bound ssDNA. (*C*) A DNA docking model was constructed by superimposing the *E. coli* SSB–DNA complex structure onto the gp2.5 OB-fold. The DNA from the *E. coli* SSB structure overlies the conserved aromatic residues (green) described in panel *B* and it is near a series of conserved basic residues (blue; cf. Fig. 4). The DNA is closely apposed to these residues without significant steric clashes. All ribbon diagrams and the electron density shown in Fig. 2 were made with the program SETOR (44) and protein surfaces (Figs. 1*C* and 3*A*) were made with the program GRASP (45).

Table 1. Crystallographic data and refinement statistics

Spacegroup	$P2_12_12_1$		
Unit cell, Å	a = 68.1, b = 71.8, c = 82.2		
Molecules/a.u.	2		
	Native	SeMet	Hg
Wavelength, Å	0.9350	0.9792	1.0073
Max resolution, Å	1.9	2.3	2.8
Completeness, %	98.6	97.5	96.5
Redundancy	5	5	4
I/σ	17	18	18
R_{merge}, %*	5.0	7.0	4.6
Number of sites	—	4	6
Resolution range, Å		25–1.9	
R factor†		21.4%	
R_{free}		26.8%	
Rmsd bond lengths, Å		0.012	
Rmsd bond angles, °		1.78	

*$R_{merge} = \Sigma|I - \langle I\rangle|/\Sigma I$, where I is the observed intensity and $\langle I\rangle$ is the average intensity.

†*R* factor $= \Sigma||F_o| - |F_c||/\Sigma|F_o|$. R_{free} is the same as *R*, but calculated with 10% of reflections that were never used in crystallographic refinement.

a long α-helix on one end of the barrel (Fig. 1). The core of the β-barrel is filled with hydrophobic residues that exclude solvent, and its surface is relatively smooth with a prominent groove formed by the extensions of the β45 loop and the β3–3a loop (Fig. 1*A*). The C-terminal 28 residues of gp2.5Δ26C extend away from the OB barrel on the side opposite the prominent groove (Figs. 1 and 3). We presume that the additional 26 residues not present in the truncated gp2.5Δ26C protein would extend from the C-terminal tail seen in the crystal structure (Fig. 1*A*). The capping helix (αA) and a short β-strand (β2a) are located in the segment connecting strands β2 and β3 of the OB-fold. αA packs against strands β1 and β3, occluding the interior of the β-barrel (Fig. 1*A*). Other ssDNA binding proteins have similar OB-folds (20–22, 37), but they lack a helix in the β23 loop analogous to αA and instead have a short α-helix that packs against strands β2 and β4 on the opposite end of the β-barrel (Fig. 3; note the

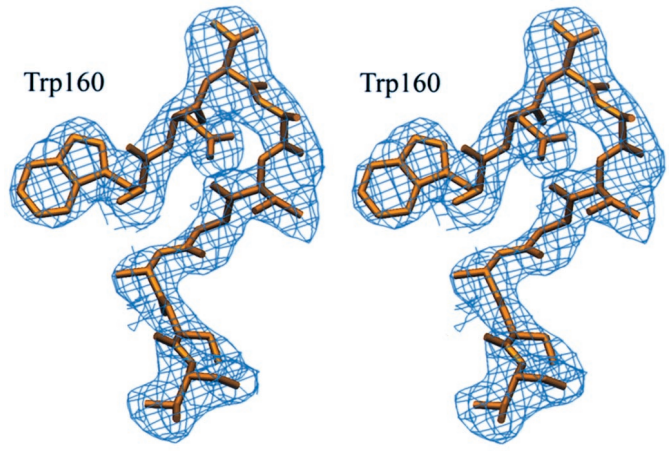

Fig. 2. Electron density from the gp2.5 crystal structure (stereo figure). Crystallographic phase information was obtained from selenium and mercury single-wavelength anomolous diffraction (SAD) phasing experiments, and subsequently improved by density modification using the program RESOLVE (28). The region around Trp160 in the β45 loop is representative of the quality of the modified experimental electron density for most of the backbone and side chain atoms of both gp2.5 subunits in the asymmetric unit.

BIOCHEMISTRY

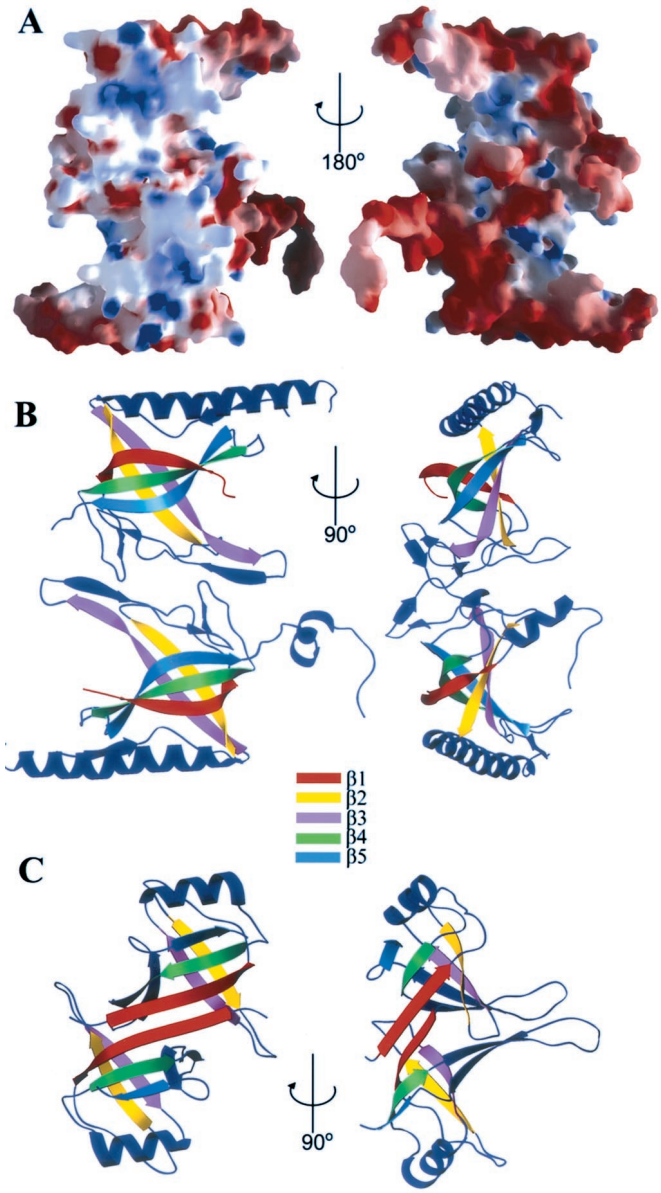

Fig. 3. Different subunit orientations in gp2.5 and *E. coli* SSB protein dimers. (*A*) The gp2.5 dimer has a distinctive electrostatic surface (calculated with GRASP; ref. 45) with one face consisting of neutral and positively charged residues (*Left*, blue) and the opposite face consisting of predominately acidic residues (*Right*, red). The proposed DNA binding surface is located in the positively charged cleft (cf. Fig. 1). (*B*) The subunits of the gp2.5 dimer pack together by an end-to-end interaction of the β-barrels involving the β23 loop, β3, and the segment between β3 and β4. The subunits of the *E. coli* SSB protein dimer (*C*) pack together by using the opposite end of the OB barrel, predominately through β1 and the connecting segment between β4 and β5. This mode of interaction is precluded by helix αA of gp2.5.

different orientations of strands β1–β5 for gp2.5 and *E. coli* SSB protein).

We believe that the crystallographic dimer of gp2.5Δ26C (Fig. 3 *A* and *B*) corresponds to the native gp2.5 dimer (2) for the following reasons. The interface between subunits of the dimer, consisting mainly of the β23 loop and the segment between β3 and β4 (Fig. 3), buries $\approx$750 Å^2 of surface area from each monomer. This substantial interface is formed predominately by nonpolar residues, although salt bridges between Glu-138 and Asn-126′ (and the reciprocal pairing between Asn-126 and Glu-138′) cross the dimer interface. A noncrystallographic 2-fold axis relates the two subunits of the dimer so that αA is located at opposite ends of the dimer and the proposed DNA binding surface (see below and Fig. 3) of both subunits is unobstructed. Although the interactions between subunits of the crystallographic dimer are rather extensive, they are not sufficient for dimerization of gp2.5Δ26C in solution. The C-terminal acidic segment of gp2.5 is required for the formation of stable dimers (6). An extended conformation of the C-terminal tail, suggested by the orientation of the adjoining residues in the gp2.5Δ26C structure, could stabilize the dimer through a domain-swapping interaction involving electrostatic interactions between acidic residues of the C-terminal tail and basic residues on the surface of the OB-fold.

Although other ssDNA binding proteins bind to ssDNA as oligomers, the available evidence is consistent with gp2.5 monomers or stable dimers binding to DNA with no additional protein oligomerization on DNA (2). *E. coli* SSB protein and the human mitochondrial SSB protein are stable tetramers, consisting of a dimer of dimers with D_2 symmetry. For both proteins the dimers associate in an end-to-end arrangement of the OB-fold with β-strands 1, 4, and 5 of each subunit packing together in an extended six-stranded antiparallel β-sheet (Fig. 3; refs. 22 and 38). This extended β-sheet of the SSB protein dimers is relatively flat, allowing two dimers to pack closely together and form the active protein tetramer (Fig. 3*C*). Helix αA of gp2.5 prevents dimerization in this orientation, and the subunits of the gp2.5 dimer pack together by using the opposite end of the OB barrel (Fig. 3*B*). Higher order oligomers of gp2.5 have not been observed. The β3′ strand and β12 loop protruding from the surface of gp2.5 prevent the dimer from packing together in a tetramer like the *E. coli* and mitochondrial SSB proteins.

The C-terminal tail of each subunit in the gp2.5 dimer extends away from the β-barrel of the OB-fold and into solution (Fig. 3). Each subunit of the gp2.5 dimer is in a different crystal packing environment. The C terminus of one subunit packs against a neighboring protomer in the crystal, and this crystal packing interaction presumably stabilizes helix αB, which is disordered in the other subunit of the gp2.5 dimer. The extended orientation and disorder observed at the C terminus of gp2.5 suggest that this segment and the additional twenty-six C-terminal residues that are deleted from gp2.5Δ26C are flexible and could adopt different conformations in gp2.5 complexed to DNA or interacting with other replication proteins. The location and proposed flexibility of the C-terminal region suggest a model in which the C-terminal tails stabilize the gp2.5 dimer (7) by a domain swapping interaction (39) across the dimer interface. A likely binding site for the highly acidic C-terminal segment is the basic region of gp2.5 that is the probable DNA binding surface (see below). Fifteen of twenty-one amino acids at the C terminus of gp2.5 are acidic residues and two residues in this short segment are aromatic. This combination of negative charge and aromatic groups is a seemingly ideal peptide mimic of ssDNA. This model predicts that binding of the C-terminal segment to the DNA binding surfaces of subunits in the gp2.5 dimer would interfere with binding to DNA, necessitating dissolution of the dimer and/or release of the tails before DNA can bind. Consistent with this proposal, a truncated gp2.5 protein lacking 21 residues at its C terminus binds to ssDNA with higher affinity than wild-type gp2.5 and has a slower rate of dissociation from DNA (J.M.S. and C.C.R., unpublished results).

Interactions of the C-terminal tail with the DNA binding site of gp2.5 could modulate its affinity for ssDNA, as was proposed for the acidic C-terminal tail of the T4 gene 32 single-stranded binding protein (40). The binding of ssDNA might act as a switch that displaces the tail so it is free to interact with the other proteins of the T7 replication complex. Conversely, the displacement of the C-terminal tail from the DNA binding surface of

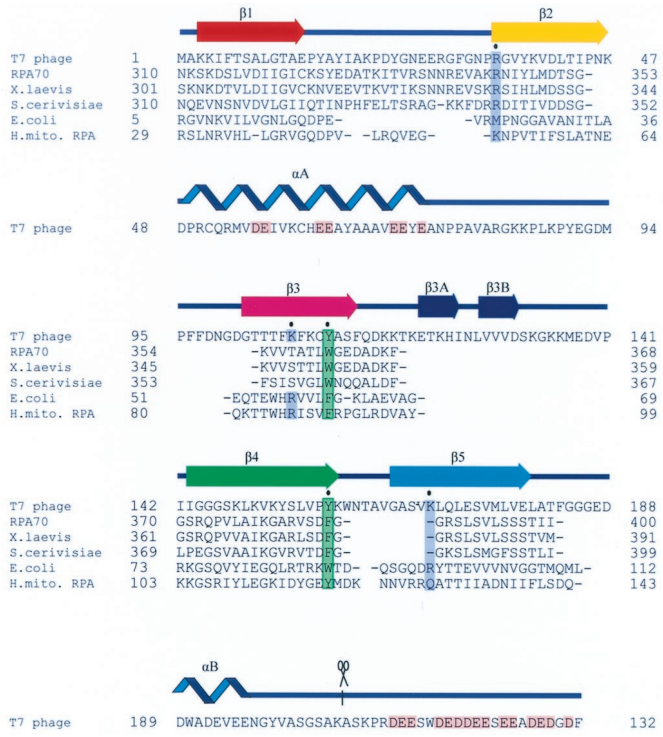

Fig. 4. Structure-based alignment of ssDNA binding protein sequences. The structures of T7 gp2.5, *E. coli* SSB protein, and one OB-domain of the human RPA70 protein were aligned (35) by using the conserved β-strands of the OB-fold. The sequences of homologous ssDNA binding proteins were then added to the aligned sequences of the structurally homologous residues. The secondary structure of gp2.5 is shown above the sequences. Although these proteins have conserved OB-folds, there is virtually no sequence conservation. The most striking feature is the alignment of aromatic residues at the end of strands 3 and 4 (green; cf. Fig. 1*C*). Structurally conserved basic residues that interact with DNA are indicated by the blue boxes (cf. Fig. 1*C*). Acidic residues contributing negative charge to the large capping helix (αA) and C terminus of gp2.5 are indicated in pink (cf. Fig. 3*A*).

gp2.5 could increase its affinity for binding to T7 DNA polymerase or T7 primase-helicase. The sequestration of the C-terminal tail in the gp2.5 dimer could be a "neutralizing" interaction that discourages binding to other proteins until gp2.5 is bound to DNA. The C-terminal tail of gp2.5 is sensitive to proteolytic digestion when gp2.5 is bound to DNA, suggesting that it is exposed in the DNA complex and available for other interactions. The model for a DNA-mediated switch in the orientation of the C-terminal tail, which is based on the crystal structure of the gp2.5 dimer, can be explored through competitive binding measurements using C-terminal peptide ligands and spectroscopic studies of the dimer.

A Model for Binding to ssDNA. The sequences of ssDNA binding proteins are homologous only for proteins from closely related organisms, and conserved sequence motifs are not readily apparent for the larger group of proteins. Despite their highly divergent sequences, proteins from each group have very similar core structures consisting of a five-stranded OB-fold. We therefore aligned the sequences of T7 gp2.5, *E. coli* SSB protein, and human RPA70 by superimposing the structures of their OB-folds with the program DALI (34, 35) to create an amino acid alignment that is based on structural homology (Fig. 4). The amino acids sequences of additional ssDNA binding proteins from *Xenopus laevis*, *Saccharomyces cerevisiae*, and human mitochondrial SSB protein were then aligned with the sequences of the founding members of the multiple alignment (Fig. 4). The structure of T4 gene 32 protein (23) is significantly different from the rest of the group (DALI Z-score = 0.7), so it was omitted from this alignment. The structure-based sequence alignment shows few amino acid similarities among these proteins, even for residues known to contact ssDNA. The exceptions are two aromatic residues (located near the C-terminal ends of β3 and β4, and lying within the prominent groove on the surface of the OB-fold; Figs. 1*B* and 4) that form part of the DNA binding surface of *E. coli* SSB protein (18) and the human RPA70 subunit (37). ssDNA binds in different orientations to the OB-folds of *E. coli* SSB protein (18) and RPA70 (37), except for three nucleotides that stack between the conserved aromatic residues described above (Figs. 1*B* and 4). The three nucleotides sandwiched between these conserved aromatic residues are flanked by the β3 strand on one side and strands β4 and β5 on the other side. Additional DNA interactions are contributed by residues within the connecting loops protruding from the OB barrel. These aromatic residues, together with the adjacent β-strands and connecting loops, constitute a trinucleotide binding motif that is conserved in gp2.5, *E. coli* SSB protein, and human RPA70.

The proposed DNA binding cleft of gp2.5 contains a mixture of nonpolar residues that could stack against the bases of DNA, punctuated with clusters of basic residues that could interact with the DNA backbone. A series of positively charged residues (Lys3, Arg35, Lys39, Lys107, Lys109, Lys150, Lys152, and Lys169) flank the nonpolar residues of trinucleotide binding motif of gp2.5 at positions where they could interact with DNA. Three of these positively charged residues (Lys39, Lys107, and Lys169) are superimposable on basic residues of *E. coli* SSB protein. Most of the remaining surface of the OB-fold of gp2.5 is negatively charged (Fig. 3*A*), further reinforcing this proposal for the location of the ssDNA binding surface.

We have modeled the proposed interaction of gp2.5 with DNA by superimposing the crystal structure of the *E. coli* SSB protein–DNA complex with six nucleotides (18) on the crystal structure of gp2.5 (Fig. 1*C*). The superposition of the OB-folds results in a reasonable docking of the DNA on the gp2.5 structure, with favorable electrostatic interactions and no severe steric clashes. The DNA docking model does not yield any compelling predictions about DNA interactions outside of the trinucleotide binding cleft of gp2.5, but it is possible that ssDNA might wrap around the barrel of the OB-fold, in a manner similar to that observed with the *E. coli* SSB protein. A stripe of nonpolar and positively charged residues flanking both sides of the trinucleotide binding cleft could mediate these interactions with DNA (Fig. 3*A*). It is still unclear, however, whether gp2.5 binds to DNA as a stable dimer or as a monomer. If the DNA docking model described above is applied to the gp2.5 dimer (Fig. 3), the ends of the bound DNAs cannot be joined to wrap around the dimer because the DNAs are oriented in opposite directions. This may suggest that gp2.5 binds to ssDNA as a monomer.

The crystal structure of bacteriophage T7 gp2.5 has revealed an OB-fold that is present in a large group of highly divergent ssDNA binding proteins. The structure of gp2.5Δ26C suggests a possible mechanism of protein dimerization in which the C-terminal tails of adjacent subunits interact *in trans* with the DNA binding surface of the apposing subunit. This model for dimerization of full length gp2.5, based on the locations of the C-termini of subunits of the gp2.5Δ26C dimer, suggests a means of modulating DNA binding affinity and protein–protein interactions by repositioning the C-terminal DNA in response to DNA binding. Two aromatic residues that stack against the bases of DNA in the *E. coli* SSB protein and RPA70 structures are conserved in the gp2.5 structure, suggestive of a similar mode of DNA binding. The favorable electrostatic surface potential of this region also supports this model for DNA binding interactions. ssDNA binding proteins play critical roles in DNA repli-

cation and recombination, and the structure of gp2.5 provides many insights into these biological activities. Together with the additional structures of T7 DNA polymerase (41) and the T7 helicase (42, 43) we can begin to piece together a high resolution picture of a DNA replication system.

We are appreciative of the expert assistance provided by the staff of beamline X-25 of the National Synchrotron Light Source (Upton, NY) and the staff of station A1 of the Macromolecular Diffraction facility at the Cornell High Energy Synchrotron Source (Ithaca, NY). We thank Michael Sawaya, Eric Toth, and Lisa Rezende for discussions and critical reading of the manuscript. This work was supported by grants from the National Institutes of Health (to T.E. and C.C.R.). T.H. was supported by a postdoctoral fellowship from the American Cancer Society. D.S.W. was supported by a postdoctoral fellowship from the National Institutes of Health.

1. Chase, J. W. & Williams, K. R. (1986) *Annu. Rev. Biochem.* **55,** 103–136.
2. Kim, Y. T., Tabor, S., Bortner, C., Griffith, J. D. & Richardson, C. C. (1992) *J. Biol. Chem.* **267,** 15022–15031.
3. Kong, D. & Richardson, C. C. (1996) *EMBO J.* **15,** 2010–2019.
4. Kong, D., Nossal, N. G. & Richardson, C. C. (1997) *J. Biol. Chem.* **272,** 8380–8387.
5. Kong, D., Griffith, J. D. & Richardson, C. C. (1997) *Proc. Natl. Acad. Sci. USA* **94,** 2987–2992.
6. Kim, Y. T. & Richardson, C. C. (1994) *J. Biol. Chem.* **269,** 5270–5278.
7. Kong, D. & Richardson, C. C. (1998) *J. Biol. Chem.* **273,** 6556–6564.
8. Lohman, T. M. & Ferrari, M. E. (1994) *Annu. Rev. Biochem.* **63,** 527–570.
9. Kim, Y. T., Tabor, S., Churchich, J. E. & Richardson, C. C. (1992) *J. Biol. Chem.* **267,** 15032–15040.
10. Nakai, H. & Richardson, C. C. (1988) *J. Biol. Chem.* **263,** 9831–9839.
11. Lee, J., Chastain, P. D., Kusakabe, T., Griffith, J. D. & Richardson, C. C. (1998) *Mol. Cell* **1,** 1001–1010.
12. Nakai, H. & Richardson, C. C. (1988) *J. Biol. Chem.* **263,** 9818–9830.
13. Kim, Y. T. & Richardson, C. C. (1993) *Proc. Natl. Acad. Sci. USA* **90,** 10173–10177.
14. Prasad, B. V. & Chiu, W. (1987) *J. Mol. Biol.* **193,** 579–584.
15. Bujalowski, W. & Lohman, T. M. (1986) *Biochemistry* **25,** 7799–7802.
16. Kowalczykowski, S. C., Lonberg, N., Newport, J. W., Paul, L. S. & von Hippel, P. H. (1980) *Biophys. J.* **32,** 403–418.
17. Kowalczykowski, S. C., Lonberg, N., Newport, J. W. & von Hippel, P. H. (1981) *J. Mol. Biol.* **145,** 75–104.
18. Raghunathan, S., Kozlov, A. G., Lohman, T. M. & Waksman, G. (2000) *Nat. Struct. Biol.* **7,** 648–652.
19. Murzin, A. G. (1993) *EMBO J.* **12,** 861–867.
20. Bochkarev, A., Bochkareva, E., Frappier, L. & Edwards, A. M. (1999) *EMBO J.* **18,** 4498–4504.
21. Yang, C., Curth, U., Urbanke, C. & Kang, C. (1997) *Nat. Struct. Biol.* **4,** 153–157.
22. Raghunathan, S., Ricard, C. S., Lohman, T. M. & Waksman, G. (1997) *Proc. Natl. Acad. Sci. USA* **94,** 6652–6657.
23. Shamoo, Y., Friedman, A. M., Parsons, M. R., Konigsberg, W. H. & Steitz, T. A. (1995) *Nature (London)* **376,** 362–366.
24. Van Duyne, G. D., Standaert, R. F., Karplus, P. A., Schreiber, S. L. & Clardy, J. (1993) *J. Mol. Biol.* **229,** 105–124.
25. Doublie, S. (1997) *Methods Enzymol.* **276,** 523–530.
26. Otwinowski, Z. & Minor, W. (1997) *Methods Enzymol.* **276,** 307–326.
27. Bailey, S. (1994) *Acta Crystallogr. D* **50,**760–763.
28. Terwilliger, T. C. (2000) *Acta Crystallogr. D* **56,** 965–972.
29. Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard (1991) *Acta Crystallogr. A* **47,** 110–119.
30. Brunger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., *et al.* (1998) *Acta Crystallogr. D* **54,** 905–921.
31. Brunger, A. T. (1992) *Nature (London)* **355,** 472–475.
32. Laskowski, R. A., McArthur, M. W., Moss, D. S. & Thornton, J. M. (1993) *J. Appl. Crystallogr.* **26,** 283–291.
33. Colovos, C. & Yeates, T. O. (1993) *Protein Sci.* **2,** 1511–1519.
34. Holm, L. & Sander, C. (1995) *Trends Biochem. Sci.* **20,** 478–480.
35. Holm, L. & Sander, C. (1996) *Methods Enzymol.* **266,** 653–662.
36. Abrahams, J. P. & Leslie, A. G. W. (1996.) *Acta Crystallogr. D* **52,** 30–42.
37. Bochkarev, A., Pfuetzner, R. A., Edwards, A. M. & Frappier, L. (1997) *Nature (London)* **385,** 176–181.
38. Webster, G., Genschel, J., Curth, U., Urbanke, C., Kang, C. & Hilgenfeld, R. (1997) *FEBS Lett.* **411,** 313–316.
39. Bennett, M. J., Schlunegger, M. P. & Eisenberg, D. (1995) *Protein Sci.* **4,** 2455–2468.
40. Wu, M., Flynn, E. K. & Karpel, R. L. (1999) *J. Mol. Biol.* **286,** 1107–1121.
41. Doublie, S., Tabor, S., Long, A. M., Richardson, C. C. & Ellenberger, T. (1998) *Nature (London)* **391,** 251–258.
42. Sawaya, M. R., Guo, S., Tabor, S., Richardson, C. C. & Ellenberger, T. (1999) *Cell* **99,** 167–177.
43. Singleton, M. R., Sawaya, M. R., Ellenberger, T. & Wigley, D. B. (2000) *Cell* **101,** 589–600.
44. Evans, S. V. (1993) *J. Mol. Graph.* **11,** 134–138.
45. Nicholls, A., Sharp, K. A. & Honig, B. (1991) *Proteins* **11,** 281–296.