Supporting Information

Zhao et al. 10.1073/pnas.0908569107

SI Text

Protein Expression and Purification. The coding region of gp1 was PCR-amplified from the Sf6 genomic DNA using the following primers: 5′- CCA CAT ATG GCG ACT GAA CCA AAA G-3′, which contains a Nde1 site followed by the N-terminal coding sequence of gp1, and 5′- GGC CTC GAG TCA TTT TCC GAA TAG AGT GCT C -3′, which harbors the C-terminus of gp1, a stop codon and a Xho1 site. The PCR product was digested with appropriate restriction enzymes, gel purified, and cloned into the same sites of the vector pET28b (Novagen). The resulting construct contained the entire coding sequence of gp1 with an N-terminal His-tag and a thrombin cleavage site before the gp1 sequence. Plasmids was transformed in *E. coli* strain BL21(DE3) plysS (Novagen). The strain was grown at 30 ºC in LB media supplemented with 30 μg/ml Kanamycin. The gp1 expression was induced at an OD600 of 0.5 by the addition of 1mM IPTG and shaken for three and half hours. Cells were harvested by centrifugation (10 min, 6,000 g), washed in cold buffer A (20 mM Tris-HCl pH8.5, 500 mM NaCl, 10 mM β-mercaptoethanol, 30 mM imidazole) and frozen at -20 °C. The cells were thawed at room temperature, and 1 mM PMSF (phenylmethylsulfonyl fluoride) was added; the suspension was placed on ice and lysed by three passages through a French press. The sample was centrifuged for 60 min at 13,000 rpm (rotor JA-20, Beckman). The supernatant was loaded on 2 mL resin Ni-NTA column balanced with buffer A, washed with buffer A and eluted with buffer A and 50 mM, 100, 200, and 500 mM imidazole. Most gp1 was eluted with 200 mM imidazole, and the fractions of 200 and 500 mM imidazole elution were pooled and loaded on size exclusion chromatography (Sephacryl S-300, GE healthcare). The single gp1 peak showed an estimated molecular weight of approximately 140 kDa and was concentrated with Ultrafree (Milipore, MW cutoff 100,000) concentrator to approximately 10 mg/mL and stored at 4 °C. The typical yield was 20 mg of purified protein per liter of bacterial culture. The N-terminal His-tag was removed by incubation of a solution of 2 U of thrombin (Sigma) per milligram of gp1 for 20 h at 4 °C. Prior to crystallization, the thrombin was removed by *p*-aminobenzamidine-agarose (Sigma) and the cleaved gp1 subunits were isolated from His-tag and uncleaved protein by Ni-affinity column chromatography. The same purification procedures were followed for Se-methionine-substituted gp1, except that the strain B834 (DE3)plysS was used.

Site-directed mutagenesis was conducted upon the plasmid pET28b encompassing the gp1 gene so that the lysine residue 59 in gp1 was substituted with a glutamate residue (K59E). The primers (5'-TGAGGATTTTCGCGACGAGTACGCGAA-GGCAAC-3' and 5'-GTTGCCTTCGCGTACTCGTCGCGAAA-ATCCTCA-3') were used. A QuikChange mutagenesis kit (Stratagene) was used for the site-directed mutagenesis. The resultant plasmid was confirmed by DNA sequencing. The expression and purification of mutant gp1-K59E is essentially the same as the native gp1.

Electron Microscopy. The purified gp1 was diluted to 0.08 mg/mL (4.5 μM), and adsorbed for 1 min on a 300 mesh Formvar copper EM grid coated with a carbon film, followed by washing with deionized water for 5 s and staining with 1% uranyl acetate for 1 min. The grid was air dried, and was inspected on a 200 kV Tecnai F20 G2 transmission electron microscope equipped with a field emission gun.

Electrophoretic Mobility Shift Assay. The gp1- and gp2-coding regions (423 and 1,413 bp respectively) in the Sf6 genome are adjacent. A DNA fragment corresponding to the DNA sequence comprising the coding regions of gp1 and gp2 (1,836 bp in total) was generated by PCR using Taq DNA polymerase (New England Biolabs) and the oligonucleotides 5′-CCA CAT ATG GCG ACT GAA CCA AAA G-3′ and 5′-GGC CTC GAG TTA CCA ACC GGA GGA TGA GGG-3′ as the primers. The purified gp1 (25, 50, 75, 100, and 125 μM) was incubated with the DNA fragment (approximately 25 nM) in 12 μL reaction solution containing 10 mM Tris-HCl pH 7.8, 100 mM NaCl, 1mM DTT and 1 mM ATP at room temperature for 30 min, and was loaded on 1% (w/v) agarose gel followed by ethidium bromide staining.

Electrophoretic mobility shift assay of Sf6 gp1 with a 1230 bp nonspecific DNA (from a Gram-positive bacterium) and of Sf6 gp1-K59E mutant with gp1-gp2 coding DNA and nonspecific DNA was performed essentially in the same way as described previously.

Fitting of Molecular Structures. The coordinates of DNA-binding domains of Sf6 gp1 and phage lambda gpNu1 were aligned using MatchMaker in the program UCSF Chimera (1) and were manually adjusted to optimize superimposition of secondary structural elements.

In the aforementioned molecular fitting, the helix α1 (residues 16–26) of Sf6 gp1 was aligned with the α-helix (residues 17–27) in the winged helix-turn-helix motif of gpNu1 that was implicated in phage lambda DNA recognition (Fig. S4). This suggests that helix α1 is directly involved in gp1:DNA interaction and can insert into the major groove of the bound DNA as in the typical DNA-binding mode of many DNA-binding proteins such as transcription factors. To model the gp1:DNA binding, we used the x-ray structure of a transcription factor Spo0A in complex with its specific DNA motif (PDB code 1lq1) (2). The α-helix (residues 210–220) in Spo0A inserted into the DNA major groove. Thus, the molecule A of the Spo0A structure with bound DNA was superimposed onto the Sf6 gp1 structure by fitting the helix (residues 210–220) of Spo0A with the helix α1 of Sf6 gp1, and the resultant DNA was replaced with a standard B-form dsDNA created with the program 3DNA (3). The DNA was slightly adjusted to avoid steric hindrance with the gp1 region of residues 34–35.

1. Pettersen EF, et al. (2004) UCSF Chimera—A visualization system for exploratory research and analysis.*J Comput Chem* 25(13):1605–1612
2. Zhao H, et al. (2002) DNA complexed structure of the key transcription factor initiating development in sporulating bacteria.*Structure* 10(8):1041–1050
3. Lu XJ, Olson WK (2008) 3DNA: A versatile, integrated software system for the analysis, rebuilding and visualization of three-dimensional nucleic-acid structures.*Nat Protoc* 3 (7):1213–1227

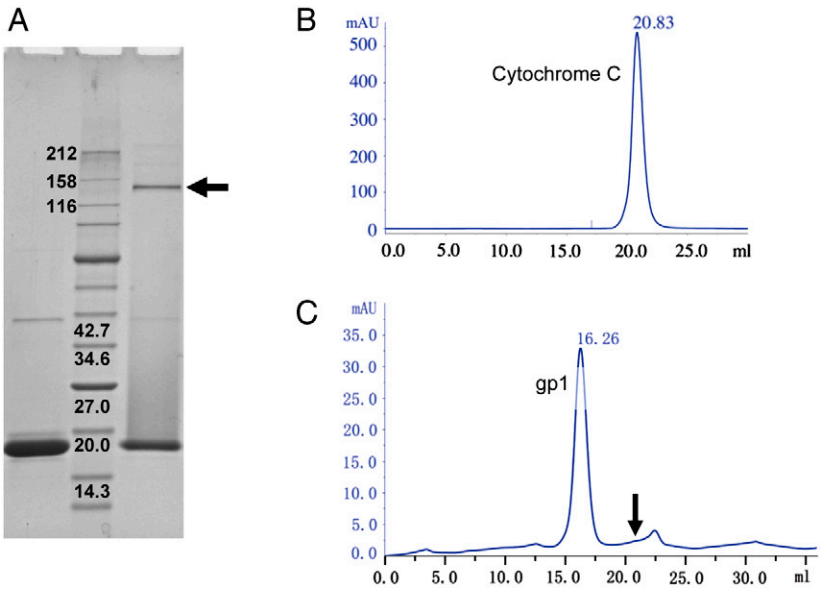

Fig. S1. Purification and Sf6 gp1. (*A*) SDS-PAGE of purified gp1. Middle lane, MW marker (kDa) (New England BioLabs). Left lane, purified His-tagged gp1 (17,734 Da). Right lane, purified gp1 without heating prior to SDS-PAGE. The band corresponding to the gp1 octamer is indicated by an arrow.

(*B*) Size exclusion chromatogram (superose 6 10/300 GL, GE Healthcare) of the molecular weight standard cytochrome C (MW=12,400Da; Sigma product C7150-1VL), showing a peak at an elution volume of 20.83 mL.

(*C*) Size exclusion chromatogram (the same column as in *B*) of gp1 showing a single major peak at an elution volume of 16.26 mL, corresponding to a molecular weight of approximately 140 kDa. No peak is present at the elution volume corresponding to the gp1 monomer (*Arrow*), indicating absence of the monomeric gp1.

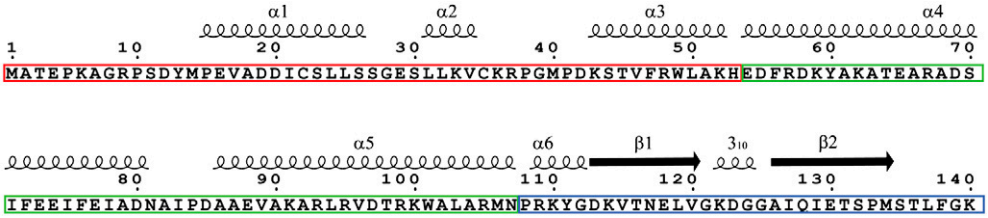

Fig. S2. The secondary structure elements of Sf6 gp1 as determined in the x-ray structure aligned with the gp1 primary sequence. The α-helices and β-strands are indicated by coils and arrows, respectively. A 3_{10} helix is also indicated. The N-terminal DNA-binding domain, the central domain, and the C-terminal domain are highlighted in red, green, and blue rectangles, respectively.

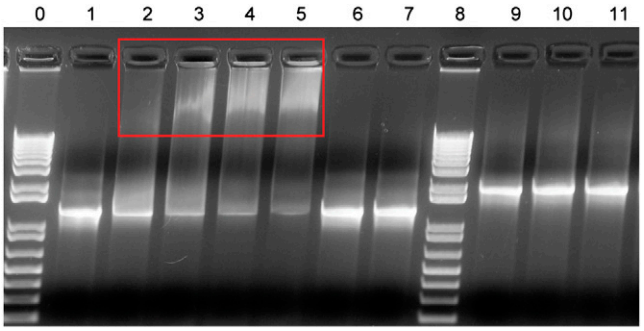

Fig. S3. Electrophoretic mobility shift assay of Sf6 gp1 with nonspecific DNA Lanes 1–5 and of Sf6 gp1-K59E mutant with gp1-gp2 coding DNA Lanes 9–11 and non-specific DNA Lanes 6–7. Lanes 0 and 8, DNA ladders. Lane 1, A 1,230 bp non-specific DNA (from a Gram-positive bacterium) alone. Lanes 2–5, nonspecific DNA incubated with 25, 50, 75, 100 μM purified Sf6 gp1 respectively, showing smeared bands (*Red Rectangle*) indicative of nonspecific binding. Lanes 6–7, nonspecific DNA incubated with 75 and 100 μM gp1-K59E mutant respectively, showing negligible DNA binding. Lane 9, gp1-gp2 coding DNA alone. Lanes 10–11, gp1-gp2 coding DNA incubated with 75 and 100 μM gp1-K59E mutant respectively, showing negligible DNA-binding.

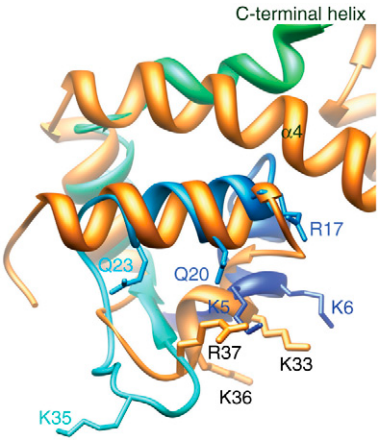

Fig. S4. Ribbon diagrams of superimposed DNA-binding domains of Sf6 gp1 and gpNu1 from phage lambda (PDB code 1j9i). Gp1 is in gold. GpNu1 is ramp-colored from the N (*Blue*) to the C terminus (*Green*). Side chains of residues involved in DNA binding of gpNu1 are shown. Also shown are side chains of corresponding positively charged residues of gp1. The α-helix C terminal to the gpNu1 DNA-binding domain is indicated with "C-terminal helix" and is close to α4 in gp1.

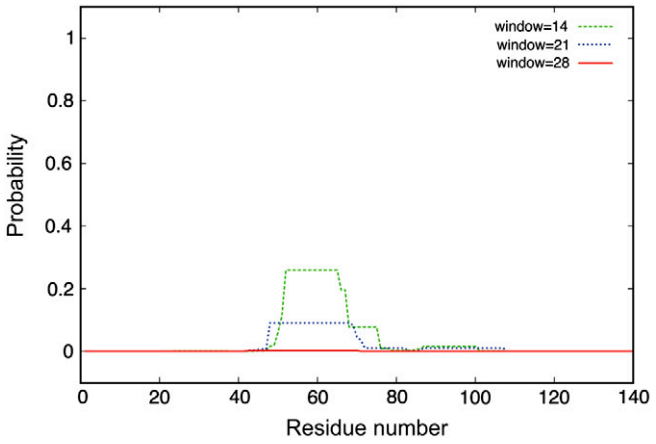

Fig. S5. Prediction with COILS (www.ch.embnet.org/software/COILS_form.html) showed a weakly probable coiled-coil region in gp1 at residues 48–68 with probability of approxinately 0.25, 0.1, and 0.0 for use of the window of 14, 21, and 28, respectively).

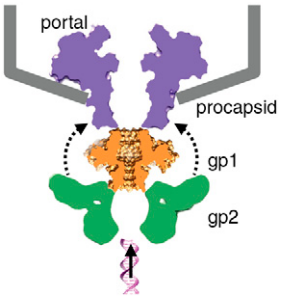

Fig. S6. An alternative model for the assembly of the Sf6 terminase and portal. Color schemes are the same as in Fig. 6. Plausible conformational change of the terminase large subunit that allows it to make contact with portal is schematically indicated with dashed arrows. The gp1 with the neck region pointing downward is shown, although the other orientation, i.e., with the neck region pointing upward, might also be possible as no direct evidence is available (see the main text for detail).

Table S1. X-ray data collection and structure refinement statistics

Data collection				
Protein	gp1 native		Se-Met gp1	
Beamline	SSRL 11-1	SSRL 7-1	SSRL 7-1	SSRL 7-1
Wavelength (Å)	1.00062	Peak 0.97910	Remote 1.03317	Inflection 0.97946
Resolution (Å)	50–1.65 (1.71–1.65)[a]	50–1.86 (1.93–1.86)	50–1.86 (1.93–1.86)	50–1.86 (1.93–1.86)
No. measurements	340,420	176,236	176,715	178,374
No. unique reflections	35,597	24,582	24,624	24,797
Completeness(%)	99.6(100.0)	98.7(99.9)	98.8(99.8)	98.7(99.5)
I/σ	63.9(5.4)	34.7(3.4)	41.0(4.0)	37.0(3.2)
*R*merge (%)[b]	4.8(46.0)	6.5(50.6)	4.9(44.2)	6.1(54.9)
Space group	*P*42(1)2	*P*42(1)2	*P*42(1)2	*P*42(1)2
Unit cell (Å)	a = 88.9, c = 72.6		a = 88.8, c = 72.5	
Structure refinement				
Resolution (Å)		20–1.65		
R_{work}/R_{free}[c]		0.22/0.25		
Number of atoms				
Protein		1,981		
Water		255		
B-factors				
Protein		26.4		
Water		41.5		
R.m.s deviations				
Bond lengths (Å)		0.011		
Bond angles (°)		1.3		
Ramachandran plot statistics				
Residues in most favoured regions		94.8%		
Residues in additional allowed regions		5.2%		
Residues in generously allowed regions		0.0%		
Residues in disallowed regions		0.0%		

[a]Values in the parentheses are for the outmost resolution shells.

[b]*R*merge $= \Sigma_{hkl}\Sigma_i|I_i(hkl) - \langle I(hkl)\rangle|/\Sigma_{hkl}\Sigma_i I_i(hkl)$, where $I_i(hkl)$ is the observed intensity of reflection hkl and $\langle I(hkl)\rangle$ is the averaged intensity of symmetry-equivalent measurements.

[c]$R_{work} = \Sigma_{hkl}||F_{obs}| - |F_{calc}||/\Sigma_{hkl}|F_{obs}|$, where F_{obs} and F_{calc} are structure factors of the observed reflections and those calculated from the refined model, respectively. R_{free} has the same formula as R_{work}, except that it was calculated against a test set of the data that was not included in the refinement.