Supporting Information

Geibel et al. 10.1073/pnas.0902910106

SI Text

The Disordered C Terminus of RepB′ Is Involved in Translocation by Type IV Secretion Systems. No electron density was found for the C-terminal 16 amino acids of full-length RepB′, ^{308}APRQRGMDRGGPDFSM^{323}, indicating disorder.

Mobilization of RSF1010 Is Accomplished by Components of Helper Plasmids Like the IncP Plasmid RP4. Mobilization protein MobA (consisting of a relaxase with C-terminally fused RepB′) can also be translocated by the Dot/Icm type IV secretion systems (T4SS) from *Legionella pneumophila* (1) and VirB/D4 T4SS from *Agrobacterium tumefaciens* (1, 2). The transport signal for T4SS resides within the C-terminal 48 amino acids of MobA/RepB′ (residues 275–323/RepB′). An RQR motif (residues 310–312/RepB′) and a net positive charge of +2 in the last 20 amino acids (RepB′ residues 304–323) (2) are supposed to promote the translocation of MobA by T4SS.

Homeodomain-Like Folds Are Not Only Found in Eukaryotes. A recent amino acid sequence interpretation showed that the C-terminal domain of protein p16.7 from bacteriophage φ29 features an amino acid sequence that is typical of homeodomains (3). Because prokaryotic RepB′ and the Iγ subdomain of phage Mu transposase also show a homeodomain-like fold (4, 5), these observations support the idea that homeodomains are not exclusively used by eukaryotes to bind dsDNA, but also by some prokaryotes. However, relevant 3D structures are missing.

Expression and Purification of RepB′. An overnight culture of *E. coli* strain SCS1 harboring the expression vector pMS470Δ8 (6) that carried the *repB′* gene of RSF1010 was grown to the midlog phase at 37 °C in LB medium. Overexpression of RepB′ was induced by the addition of 1 mM isopropylthiogalactoside. The cells were harvested after 4 h by centrifugation, resuspended in 50 mM Tris·HCl (pH 7.0), 200 mM $MgSO_4$, and 2 mM DTT and disrupted at 4 °C with a French Press. After centrifugation the supernatant contained RepB′ that was purified by FPLC chromatography using Hi-Trap heparin (GE Healthcare) and POROS 20 HS (Applied Biosystems) columns and applying a linear gradient of 0–1 M NaCl in 20 mM Tris·HCl (pH 7.0), 50 mM $MgSO_4$, and 1 mM DTT. Fractions containing RepB′ were gel-filtrated on a Superdex75 column (GE Healthcare) using 10 mM Tris·HCl (pH 7.0), 200 mM $MgSO_4$, and 1 mM DTT and concentrated (Amicon; Millipore) in the range of 7 to 14 mg/mL for crystallization. RepB′ was identified by N-terminal sequencing and MALDI-TOF mass spectrometry.

Expression and Purification of the N- and C-Terminal Domains. The proteins were overexpressed in *E. coli* strain ER2566 and purified on a HisTrap-column (GE Healthcare) applying a gradient of 0–500 mM imidazole in 200 mM $MgCl_2$, 20 mM Hepes (pH 8), and 1 mM DTT. The His_6 tags were cleaved by thrombin and removed by dialysis against the same buffer, yielding pure C-terminal domain, whereas the N-terminal domain was additionally purified as described above for RepB′.

RepB′: Crystallization, Data Collection, and Processing. Hanging drop vapor diffusion at 18 °C yielded rod-like crystals, 0.4 × 0.1 × 0.1 mm^3, from the solution of RepB′ described above supplemented with 1 mM dTTP and mixed with the same volume (1 μL) of reservoir buffer [0.3 M ammonium sulfate, 22% (wt/vol) PEG monomethyl ether 2000, 5% (vol/vol) γ-butyrolactone, and 100 mM sodium acetate pH 5.5]. Soaking crystals with OsO3 × 2 pyridine (10 mM) for 24 h afforded a heavy atom derivative.

Crystals of RepB′ were transferred to reservoir buffer supplemented with 20% (vol/vol) glycerol as cryoprotectant and flash frozen in liquid nitrogen. Data were collected at 100 K (beamline ID14–2, European Synchrotron Radiation Facility/Grenoble, with ADSC Q4 CCD detector). MAD data sets were collected at 100 K on an OsO3 derivative crystal using a MAR CCD 165-mm detector at PSF beamline BL-14.1 of Freie Universität Berlin at Berliner Elektronenspeicherring-Gesellschaft für Synchrotronstrahlung/Berlin. The images were processed and scaled with HKL2000 (7) and XDS (8) (Table S1).

RepB′: Structure Determination and Refinement. Difference Patterson methods using SHELX (9) and HKL2MAP (10) located 3 osmium sites in the crystal asymmetric unit. After MAD phasing, the solvent-flattened electron density map could be partly autotraced by RESOLVE (11). After manual model building with COOT (12) and conjugate gradient minimization with REFMAC5 (13) using TLS groups in the final restrained refinement (14), the model was completed and refined until it converged. WHATCHECK (15) and PROCHECK (16) confirmed good stereochemistry with all Φ, Ψ torsion angles in allowed areas. The model includes 89% of the 323 amino acids of full-length RepB′ (Table S1).

Catalytic Domain/*ssiA*(3′Δ13) Complex: Crystallization, Data Collection, and Processing. Using hanging drop vapor diffusion, rod-like crystals grew at 18 °C to 0.3 × 0.05 × 0.05 mm^3 from the solution of the purified catalytic domain/*ssiA*(3′Δ13) complex (see above), mixed with the same volume (1 μL) of reservoir buffer [0.2 M $(NH)_4$-citrate, pH 5 and 20% (wt/vol) PEG 3350].

The crystals were transferred to reservoir solution supplemented with 30% (vol/vol) PEG-400 as cryoprotectant and flash-frozen in liquid nitrogen. Data were collected at 100 K (beamline 14.2 of Freie Universität Berlin at Berliner Elektronenspeicherring-Gesellschaft für Synchrotronstrahlung/Berlin equipped with MAR CCD 165-mm detector). The diffraction improved by 0.7 Å when the cryo-stream was repeatedly interrupted for a few seconds until convergence at 2.7-Å resolution.

Catalytic Domain/*ssiA*(3′Δ13): Structure Determination and Refinement. The phase problem was solved by molecular replacement (17) using the structure of the catalytic domain (residues 4–205) as the search model. *ssiA*(3′Δ13) was manually modeled into the electron density with COOT (12), using a model of B-DNA for the double helical stem. The model was refined until convergence after conjugate gradient minimization with REFMAC5 (13) and verified by WHATCHECK (15) and PROCHECK (16). All Φ, Ψ torsion angles were in the allowed range (Table S4).

EMSA with RepB′ and Catalytic Domain. Three different ssDNA molecules were used in EMSA (nucleotides in the hairpin in bold). The hairpin as proposed earlier (18) (Fig. S1B) is underlined. *ssiA* DNA, intact (5′-CCTTT**CCCCCTACCCGAAGGGTGGGGG**CGCGTGTGCAGCC-3′), *ssiA*(3′Δ13), truncated 3′-tail (5′-CCTTT**CCCCCTACCCGAAGGGTGGGGG**-3′), *ssiA*(5′Δ5), truncated 5′-tail (5′- **CCCCCTACCCGAAGGGTGGGGG**CGCGTGTGCAGCC-3′).

These DNAs were synthesized and purified by TIB MOLBIOL. Their purity was confirmed by electrophoresis on a 20%

polyacrylamide gel containing 7 M urea. The reaction mixtures containing 10 μM *ssiA* DNA and 5–12.5 μM RepB′ in 50 mM $MgCl_2$, 50 mM KCl, and 20 mM Tris·HCl, pH 7.5 were incubated at 37 °C for 60 min and loaded onto a 6% native polyacrylamide gel with 0.5× Tris-acetate-EDTA (TAE) buffer. Electrophoresis was performed at 6 V/cm at 4 °C for 170 min after a prerun of 30 min under the same conditions, and DNA was stained with ethidium bromide (Fig. 7).

EMSA with Helix-Bundle Domain of RepB′. EMSA with 10 μM *ssiA* DNA or truncated *ssiA* DNA′s (see above) and 5–12.5 μM of the helix-bundle domain (residues 212–323) of RepB′ was performed as described above except (because the C-terminal domain was only soluble under high-salt conditions) with 200 mM $MgCl_2$, 20 mM Hepes (pH 8.0), polyacrylamide gel [8% polyacrylamide, 6% glycerol, 20 mM Hepes (pH 8.0), 0.1 mM EDTA], and electrophoresis buffer [20 mM Hepes (pH 8.0), 0.1 mM EDTA] (Fig. 5). Attempts to use only hairpin dsDNA (nucleotides 6–27) in EMSA with the 5-helix bundle were abandoned because the hairpin DNA was not soluble under the above conditions.

Analytical ultracentrifugation was performed following described procedures (19). Protein and DNA were dissolved in buffer containing 100 mM $MgCl_2$, 50 mM KCl, 20 mM Hepes, pH 8.0. Extinction coefficients ε (l/mol × cm) used for these studies were calculated by using ProtParam (20) from the expasy-server. The values used were ε = 29,450 for the catalytic domain, ε = 2,980 for the helix-bundle domain, ε = 32,555 for RepB′, ε = 266,190 for 27-mer *ssiA* DNA(3′Δ13), and ε = 393,490 for 40-mer *ssiA* DNA.

1. Luo ZQ, Isberg RR (2004) Multiple substrates of the *Legionella pneumophila* Dot/Icm system identified by interbacterial protein transfer. *Proc Natl Acad Sci USA* 101:841–846.
2. Vergunst AC, et al. (2005) Positive charge is an important feature of the C-terminal transport signal of the VirB/D4-translocated proteins of *Agrobacterium*. *Proc Natl Acad Sci USA* 102:832–837.
3. Munoz-Espin D, et al. (2004) Phage phi29 DNA replication organizer membrane protein p16.7 contains a coiled coil and a dimeric, homeodomain-related, functional domain. *J Biol Chem* 279:50437–50445.
4. Kuo CF, et al. (1991) DNA–protein complexes during attachment-site synapsis in Mu DNA transposition. *EMBO J* 10:1585–1591.
5. Zou AH, Leung PC, Harshey RM (1991) Transposase contacts with mu DNA ends. *J Biol Chem* 266:20476–20482.
6. Balzer D, et al. (1992) KorB protein of promiscuous plasmid RP4 recognizes inverted sequence repetitions in regions essential for conjugative plasmid transfer. *Nucleic Acids Res* 20:1851–1858.
7. Otwinowski Z, Minor W, eds (1997) *Processing of X-Ray Diffraction Data Collected in Oscillation Mode* (Academic, New York).
8. Brünger AT, et al. (1998) Crystallography and NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr D* 54:905–921.
9. Schneider TR, Sheldrick GM (2002) Substructure solution with SHELXD. *Acta Crystallogr D* 58:1772–1779.
10. Pape T, Schneider TR (2004) HKL2MAP: A graphical user interface for phasing with SHELX programs. *J Appl Crystallogr* 37:843–844.
11. Terwilliger TC (2003) Automated main-chain model building by template matching and iterative fragment extension. *Acta Crystallogr D* 59:38–44.
12. Emsley P, Cowtan K (2004) Coot: Model-building tools for molecular graphics. *Acta Crystallogr D* 60:2126–2132.
13. Murshudov GN, Vagin AA, Dodson EJ (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr D* 53:240–255.
14. Winn MD, Isupov MN, Murshudov GN (2001) Use of TLS parameters to model anisotropic displacements in macromolecular refinement. *Acta Crystallogr D* 57:122–133.
15. Hooft RW, Vriend G, Sander C, Abola EE (1996) Errors in protein structures. *Nature* 381:272.
16. Morris AL, MacArthur MW, Hutchinson EG, Thornton JM (1992) Stereochemical quality of protein structure coordinates. *Proteins* 12:345–364.
17. Storoni LC, McCoy AJ, Read RJ (2004) Likelihood-enhanced fast rotation functions. *Acta Crystallogr D* 60:432–438.
18. Honda Y, et al. (1993) Mutational analysis of the specific priming signal essential for DNA replication of the broad host-range plasmid RSF1010. *FEBS Lett* 324:67–70.
19. Behlke J, Ristau O, Schönfeld HJ (1997) Nucleotide-dependent complex formation between the *Escherichia coli* chaperonins GroEL and GroES studied under equilibrium conditions. *Biochemistry* 36:5149–5156.
20. Wilkins MR, et al. (1999) Protein identification and analysis tools in the ExPASy server. *Methods Mol Biol* 112:531–552.
21. Lin LS, Kim YJ, Meyer RJ (1987) The 20-bp, directly repeated DNA sequence of broad host range plasmid R1162 exerts incompatibility in vivo and inhibits R1162 DNA replication in vitro. *Mol Gen Genet* 208:390–397.
22. Miao DM, et al. (1995) The interaction of RepC initiator with iterons in the replication of the broad host-range plasmid RSF1010. *Nucleic Acids Res* 23:3295–3300.
23. Niedenzu T, et al. (2001) Crystal structure of the hexameric replicative helicase RepA of plasmid RSF1010. *J Mol Biol* 306:479–487.
24. Scherzinger E, et al. (1997) The RepA protein of plasmid RSF1010 is a replicative DNA helicase. *J Biol Chem* 272:30228–30236.
25. Frick DN, Richardson CC (2001) DNA primases. *Annu Rev Biochem* 70:39–80.
26. Miao DM, et al. (1993) A base-paired hairpin structure essential for the functional priming signal for DNA replication of the broad host range plasmid RSF1010. *Nucleic Acids Res* 21:4900–4903.
27. Lin LS, Meyer RJ (1987) DNA synthesis is initiated at two positions within the origin of replication of plasmid R1162. *Nucleic Acids Res* 15:8319–8331.
28. Tanaka K, et al. (1994) Functional difference between the two oppositely oriented priming signals essential for the initiation of the broad host-range plasmid RSF1010 DNA replication. *Nucleic Acids Res* 22:767–772.
29. Haring V, Scherzinger E (1989) in *Promiscuous Plasmids of Gram-Negative Bacteria*, ed Thomas CM (Academic, London), pp 95–124.
30. Billeter M, et al. (1993) Determination of the nuclear magnetic resonance solution structure of an Antennapedia homeodomain-DNA complex. *J Mol Biol* 234:1084–1093.
31. Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties, and weight matrix choice. *Nucleic Acids Res* 22:4673–4680.
32. Gouet P, Courcelle E, Stuart DI, Metoz F (1999) ESPript: Analysis of multiple sequence alignments in PostScript. *Bioinformatics* 15:305–308.
33. Rawlings DE, Tietze E (2001) Comparative biology of IncQ and IncQ-like plasmids. *Microbiol Mol Biol Rev* 65:481–496.
34. Altschul SF, et al. (1997) Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Res* 25:3389–3402.

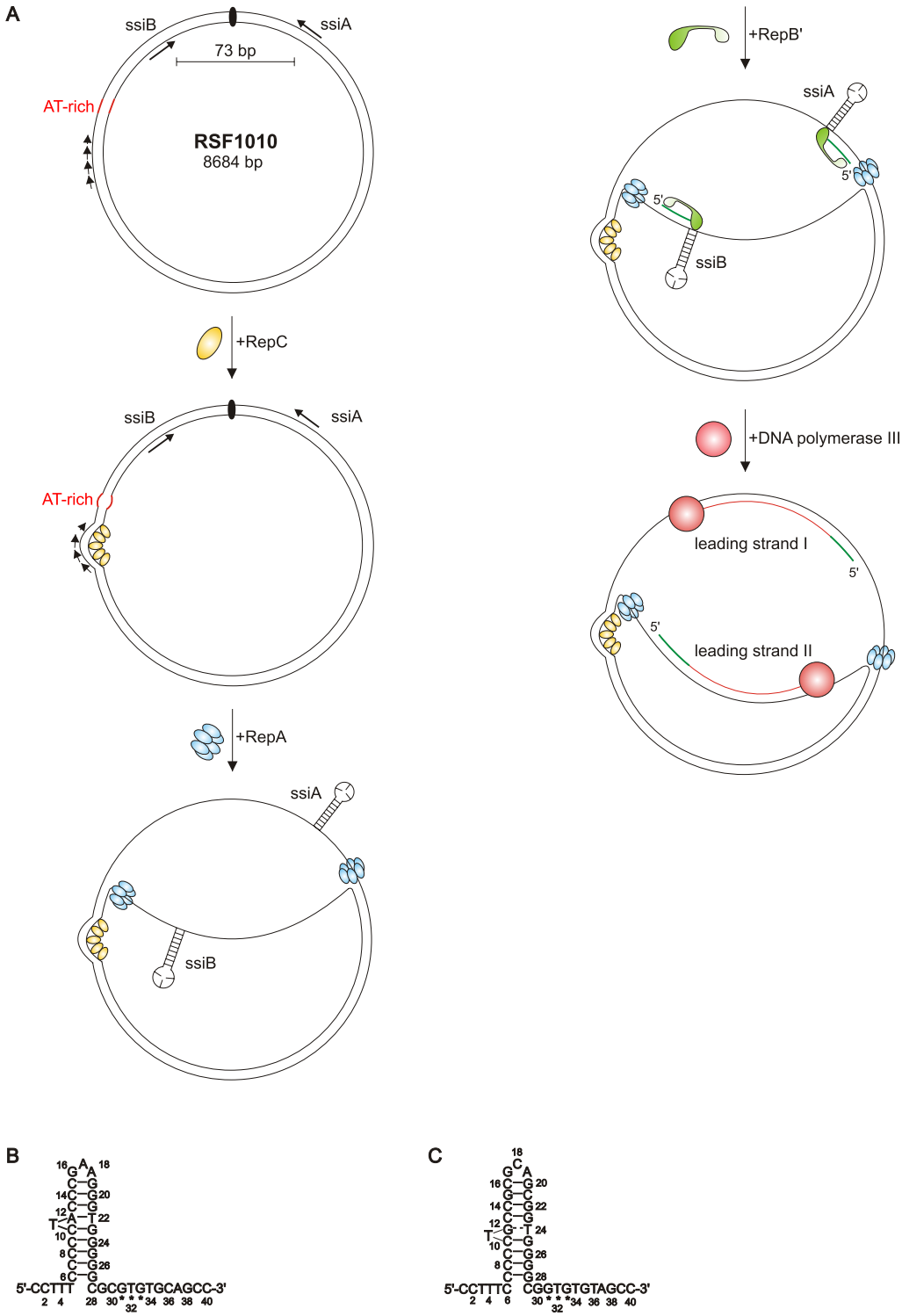


Fig. S1. (*A*) Replication of RSF1010. After initiator protein RepC binds to 3 20-bp-long direct repeats [iterons that exert plasmid incompatibility (21)] in the 396-bp-long origin of vegetative replication (*oriV*), an AT-rich site next to the iterons is opened (22) to which IncQ-specific replicative helicase RepA is loaded. RepA is a ring-shaped homohexamer (23) that invades and unwinds dsDNA with 5′ to 3′ polarity (24) by possibly channeling DNA through the central pore as described for *E. coli* DnaB helicase (25). For RSF1010 replication, primase RepB′ requires 2 specific 40-nt-long sites termed ssDNA initiators A and B, *ssiA* and *ssiB*, with highly-similar sequences that are arranged like a palindrome in the *oriV* and likely form hairpin structures (18, 26) after strand separation. RepB′ recognizes the *ssiA* and *ssiB* independently from each other and synthesizes primers that are finally extended by DNA polymerase in leading- (but not lagging-) strand replication mode in opposite directions along ss template DNA (27–29). (*B* and *C*) Sequences and hairpin structures of 40-nt-long *ssiA* and *ssiB*. The * below the sequences indicate the priming sites (27). *ssiA* (*B*) and *ssiB* (*C*) secondary structure with flipped-out T11 as proposed earlier (27) are shown.

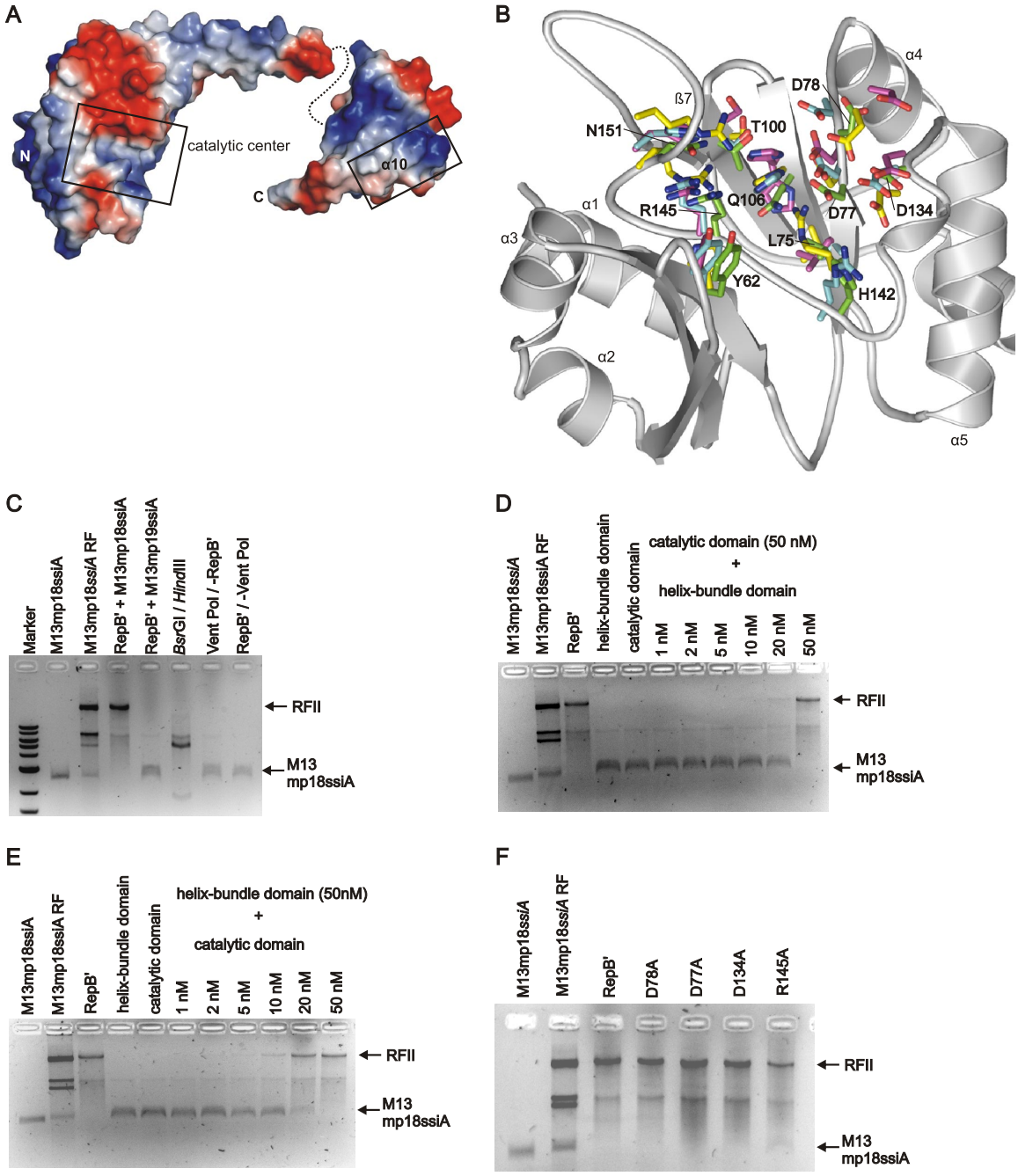


Fig. S2. (*A*) Surface electrostatic potential of RepB′. Same view as Fig. 1*B*) to show positively-charged (blue), negatively-charged (red), and neutral surface (gray). (*B*) Catalytic domain with conserved amino acids in the active center. Similar view as in Fig. 2*B*. Amino acids are presented as sticks, C atoms are colored depending on primase: RepB′ (green), Pfu primase (blue), Sso primase (magenta), and ligD polymerase domain (yellow). Oxygen is red and nitrogen is blue. The amino acids of RepB′ are denoted in 1-letter code. Equivalent amino acids of Pfu, Sso primase, and LigD-polymerase domain can be deduced by color code and Table S2. For example, the equivalent amino acids to D134 (green, RepB′) are D280 (blue, Pfu), D759 (yellow, ligD), and D235 (magenta, Sso). Amino acids of Pho primase are not shown for clarity, but occupy the same position as their equivalents in Pfu primase and share the same numbers. Pfu and Pho primases have 90% amino acid sequence similarity. (*C*) Sequence-specific primase activity of RepB′. See the legend for Fig. 3*B*. M13mp18*ssiA*, ssDNA M13 carries *ssiA*. M13mp19*ssiA*, ssDNA M13 harbors the complementary strand of *ssiA*. M13mp18*ssiA*+RepB′/M13mp19*ssiA*+RepB′, complete reaction mixtures including either M13mp18*ssiA* and RepB′ or M13mp19*ssiA* and RepB′. (*D*) Only equimolar ratio of catalytic and helix-bundle domains reconstitutes primase activity fully. See the legend for Fig. 3*B*. The concentration of the helix-bundle domain was decreased from 50 to 20, 10, 5, 2, and 1 nM at constant concentration of the catalytic domain (50 nM). Controls: RepB′, M13mp18*ssiA* RF isolated from *E. coli* and ssM13mp18*ssiA*. (*E*) Catalytic and helix-bundle domains reconstitute primase activity only when equimolar ratio is applied. See the legend for Fig. 3*B*. The concentration of the catalytic domain was decreased from 50 to 20, 10, 5, 2, and 1 nM at constant concentration of the helix-bundle domain (50 nM). Controls: RepB′, M13mp18*ssiA* RF isolated from *E. coli* and ssM13mp18*ssiA*. (*F*) Complementation assay: RepB′ variants supplemented with catalytic domain. See the legend for Fig. 3*B*. D77A, D78A, D134A, R145A, reactions with the catalytic domain (50 nM) and 50 nM of variant D77A or D78A or D134A or R145A. Controls: RepB′, M13mp18*ssiA* RF isolated from *E. coli* and ssM13mp18*ssiA*.

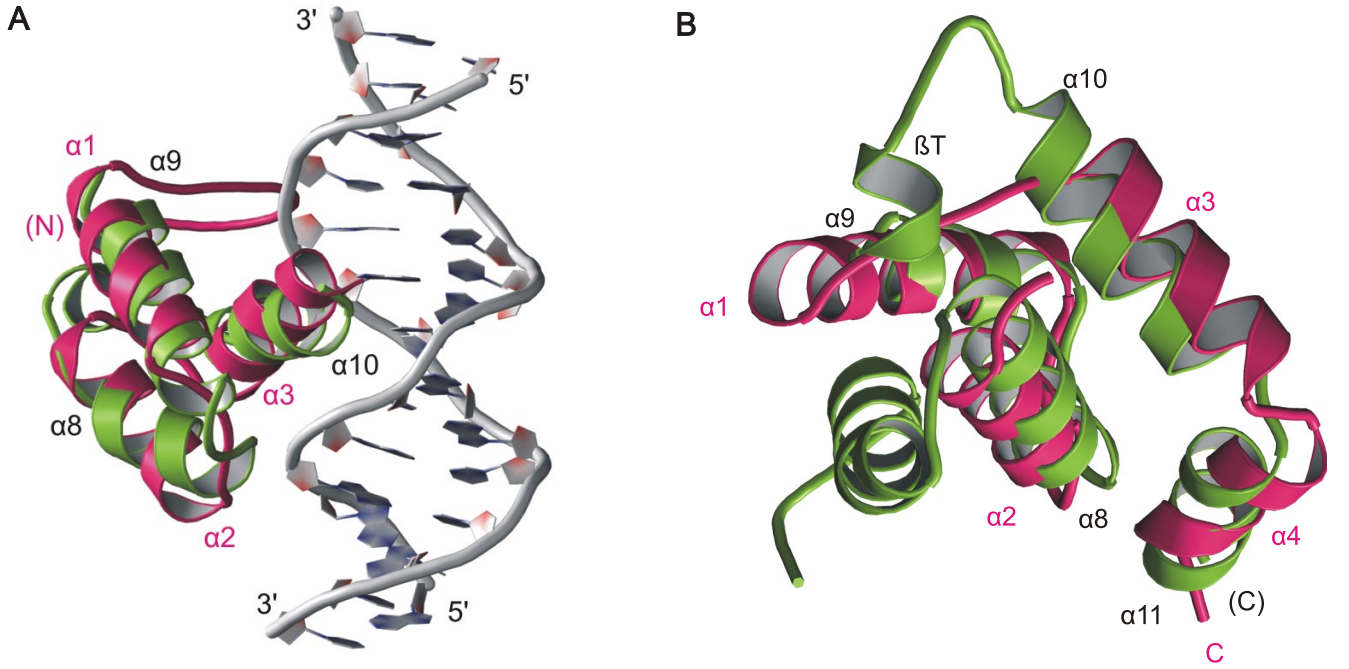


Fig. S3. (*A*) The homeodomain-like fold of the helix-bundle domain of RepB′ modeled on dsDNA. Helices α8–α10 of RepB′ (green) are superimposed on the *Antp* homeodomain from *Drosophila melanogaster* (red; α1–α3) bound to cognate DNA drawn as ribbon model (gray; Protein Data Bank ID code 1AHD) (30). The putative regognition helix α10 of RepB′ inserts well into the major groove of DNA; helices α7 and α11 of RepB′ (that would not interfere sterically when binding to dsDNA) are not shown for clarity. (*B*) The putative DNA binding parts of the helix-bundle domain of RepB′ (green) are superimposed on the phage Mu transposase Iγ homeodomain (red). Superimposition (Mu; Protein Data Bank ID code 2ezh-A) yields a Z-score of 3.3 and an rmsd of 2.9 Å for 58 C_α positions. Helices α1–α3 of Mu transposase and helices α8–α10 of RepB′ comprise the homeodomain-like fold. Additionally, helices α4 of Mu and α11 of RepB′ superimpose well.

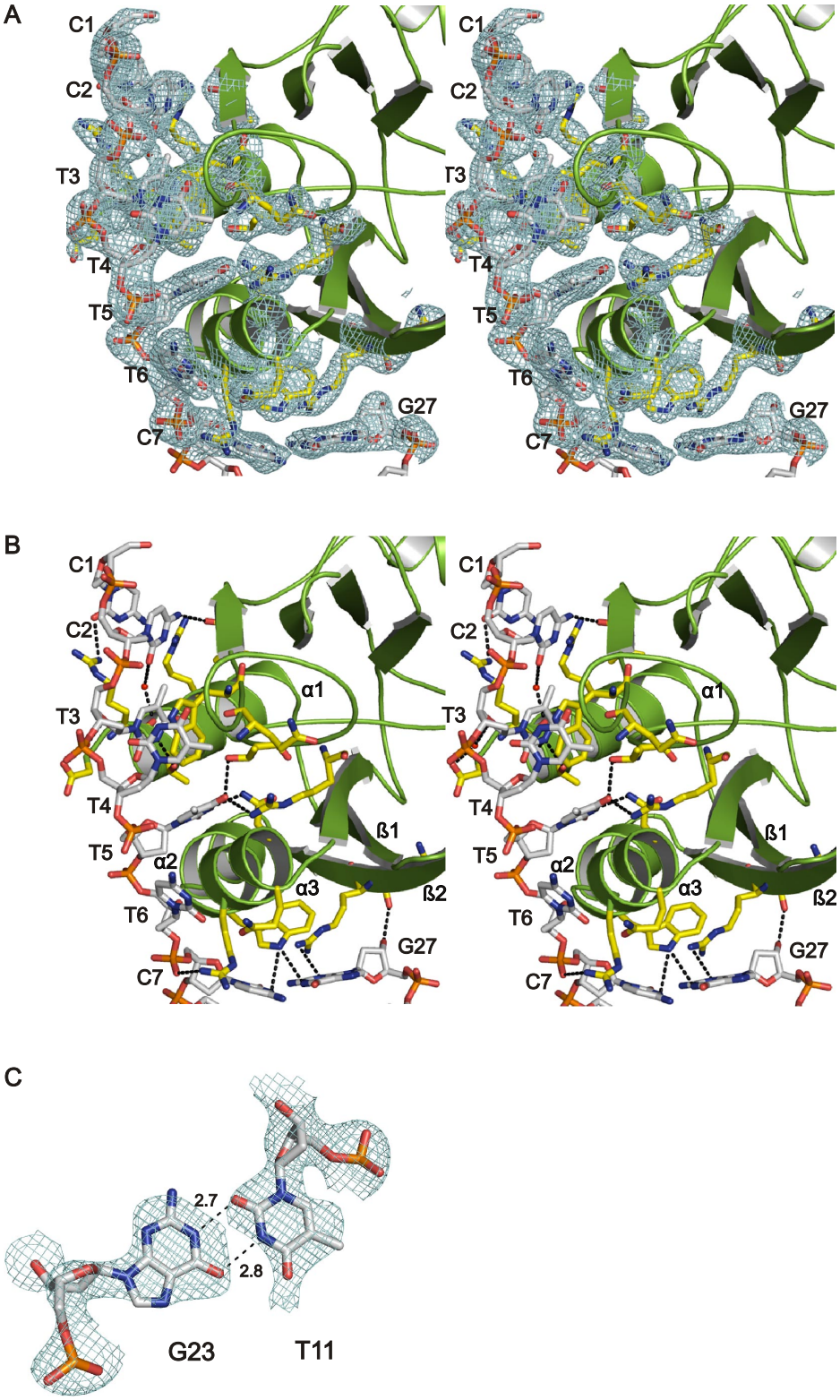


Fig. S4. (*A*) Stereoview of the binding interface of catalytic domain and *ssiA*(3′Δ13) DNA. Electron density is only shown for interacting amino acids and DNA and drawn as blue mesh (contoured with sigma level = 1.0). Bound amino acids (yellow) and DNA (gray) are presented as stick model. (*B*) Same view as in *A*, but without electron density. Hydrogen bonds are indicated by black dashed lines. (*C*) The G–T-wobble found in the hairpin helix. Bases G23 and T11 are presented as stick models (carbon gray, oxygen red, nitrogen blue). They form a wobble base pair with 2 hydrogen bonds (dashed lines, numbers give N····O distances in Å). Electron density as blue mesh at 1.0-sigma level.

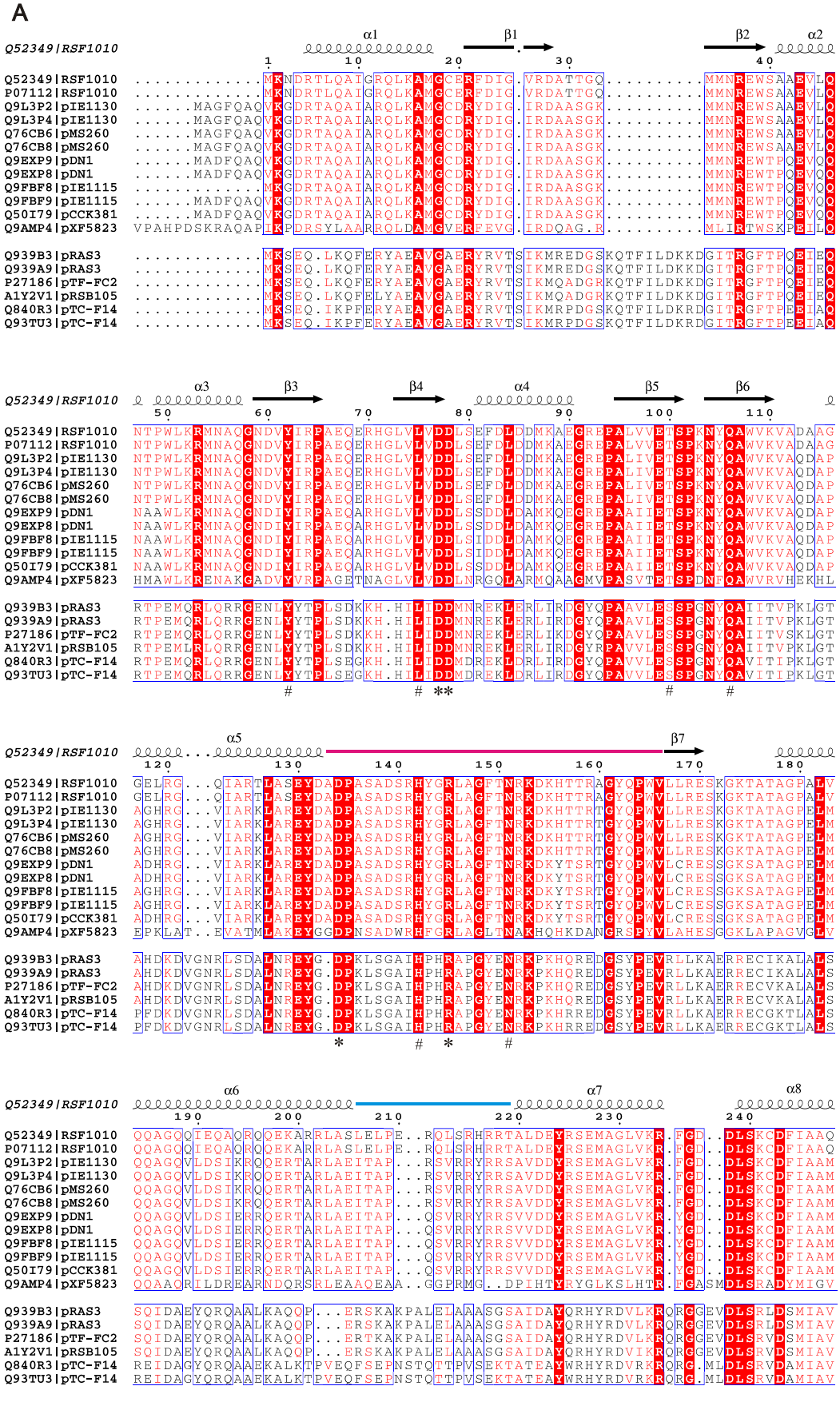
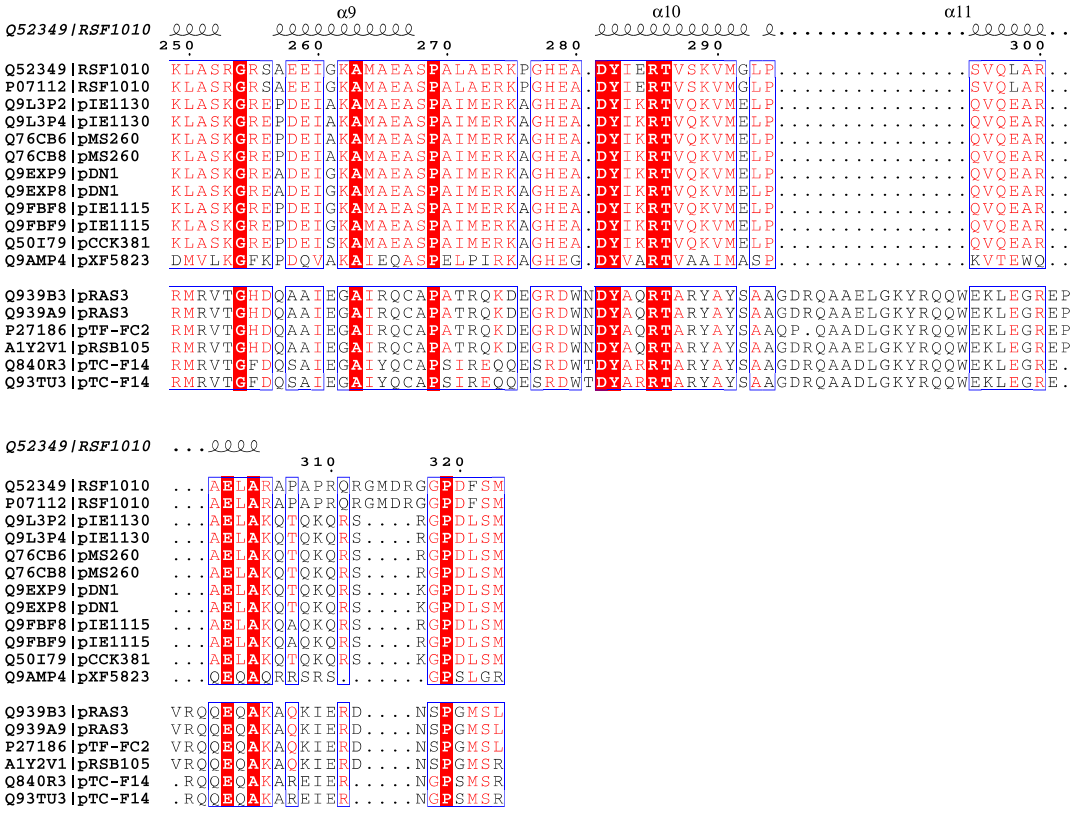

A
Q52349|RSF1010
P07112|RSF1010
Q9L3P2|pIE1130
Q9L3P4|pIE1130
Q76CB6|pMS260
Q76CB8|pMS260
Q9EXP9|pDN1
Q9EXP8|pDN1
Q9FBF8|pIE1115
Q9FBF9|pIE1115
Q50I79|pCCK381
Q9AMP4|pXF5823
Q939B3|pRAS3
Q939A9|pRAS3
P27186|pTF-FC2
A1Y2V1|pRSB105
Q840R3|pTC-F14
Q93TU3|pTC-F14

Fig. S5A


Q52349|RSF1010
α9
α10
α11
250
260
270
280
290
300
Q52349|RSF1010
P07112|RSF1010
Q9L3P2|pIE1130
Q9L3P4|pIE1130
Q76CB6|pMS260
Q76CB8|pMS260
Q9EXP9|pDN1
Q9EXP8|pDN1
Q9FBF8|pIE1115
Q9FBF9|pIE1115
Q50I79|pCCK381
Q9AMP4|pXF5823
Q939B3|pRAS3
Q939A9|pRAS3
P27186|pTF-FC2
A1Y2V1|pRSB105
Q840R3|pTC-F14
Q93TU3|pTC-F14
310
320

B

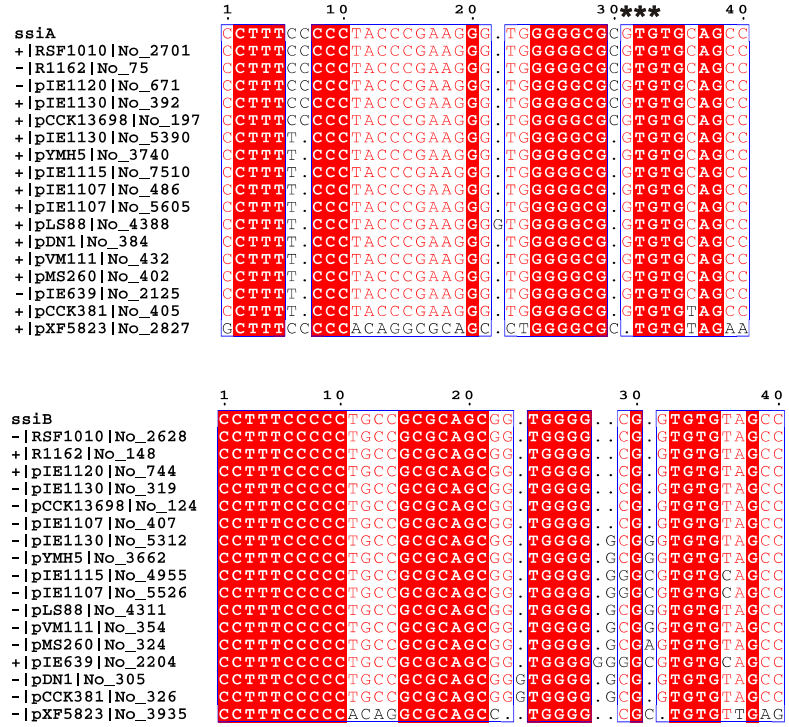

1
10
20
30***
40
ssiA
+|RSF1010|No_2701
-|R1162|No_75
-|pIE1120|No_671
+|pIE1130|No_392
+|pCCK13698|No_197
+|pIE1130|No_5390
+|pYMH5|No_3740
+|pIE1115|No_7510
+|pIE1107|No_486
+|pIE1107|No_5605
+|pLS88|No_4388
+|pDN1|No_384
+|pVM111|No_432
+|pMS260|No_402
-|pIE639|No_2125
+|pCCK381|No_405
+|pXF5823|No_2827
ssiB
-|RSF1010|No_2628
+|R1162|No_148
+|pIE1120|No_744
-|pIE1130|No_319
-|pCCK13698|No_124
-|pIE1107|No_407
-|pIE1130|No_5312
-|pYMH5|No_3662
-|pIE1115|No_4955
-|pIE1107|No_5526
-|pLS88|No_4311
-|pVM111|No_354
-|pMS260|No_324
+|pIE639|No_2204
-|pDN1|No_305
-|pCCK381|No_326
-|pXF5823|No_3935

Fig. S5B

Fig. S5. (*A*) Amino acid sequence alignment of RepB′ with similar proteins found in the Swiss-Prot/TrEMBL database. Designations of protein sequences with data bank entry code| plasmid name. Alignment was by CLUSTALW (31) and modified with ESPript (32). The secondary structure elements of RepB′ are displayed above the alignment of proteins. The red horizontal line represents the loop located in the active site (residues 133–166) and the blue line indicates the tether between helix α6 and the C-terminal domain (residues 206–219). Identical residues shared by the aligned proteins are in white letters on red background and homologous amino acids are in red letters. Strictly-conserved or homologous amino acids that are shared by the structures of RepB′, LigD-polymerase, and Sso, Pfu, and Pho primases and found within the active centeres are indicated by * (strictly conserved) or # (homologous) (see Table S2). The sequences are clearly divided into 2 groups (as previously shown in ref. 33). The larger group (top) contains plasmids RSF1010 (and near-identical R1162), pIE1115, pIE1130, and related plasmids, and the smaller group (bottom) contains plasmids pTC-F14, pTF-FC2, and others. Among the 2 groups, the amino acid identity is only 15%, but within the groups it is much higher. The larger group (top) shows >38% sequence identity and >63% homology. Omitting protein Q9AMP4, the sequence identity is >70% and the homology >93%. Omitting protein Q9AMP4, the sequence identity among the smaller group (bottom) is >76%, and the homology is >88%. It is typical for the (putative) primases compared here that they are also found as C-terminal domains fused to (putative) relaxases. These domains were also included in this sequence alignment and are marked by (+) in the 2 lists (plasmids of the top group, plasmids of the bottom group), the N-terminal parts representing relaxases are omitted for clarity. For Q9AMP4-, Q50I79-, and A1Y2V1-mobilization proteins, no related and independently expressed primases have been found in the Swiss-Prot/TrEMBL database. Plasmids of the top group include: Q52349, primase RepB′, broad-host-range IncQ plasmid RSF1010; P07112, mobilization protein A (+), broad-host-range IncQ plasmid RSF1010; Q9L3P2, putative primase, broad-host-range IncQ-like plasmid pIE1130; Q9L3P4, putative mobilization protein (+), broad-host-range IncQ-like plasmid pIE1130; Q76CB6, putative primase, plasmid pMS260; Q76CB8, mobilization protein A (+), plasmid pMS260; Q9EXP9, mobilization protein A (+), broad-host-range plasmid pDN1; Q9EXP8, replication protein B, broad-host-range plasmid pDN1; Q9FBF8, putative primase, broad-host-range plasmid IncQ-like plasmid pIE1115; Q9FBF9, putative mobilization protein (+), broad-host-range IncQ-like plasmid pIE1115; Q50I79, mobilization protein A (+), plasmid pCCK381; Q9AMP4, RepB/MobA-like protein (+), plasmid pXF5823. Plasmids of the bottom group include: Q939B3, polyprotein, plasmid pRAS3.1 (+); Q939A9, polyprotein, plasmid pRAS3.1; P27186, replicative REPB_THIFE DNA primase, broad-host-range IncQ-like plasmid pTF-FC2; A1Y2V1, mobilization protein (+), plasmid pRSB105; Q840R3, relaxase/primase-like fusion protein (+), broad-host-range IncQ-like plasmid pTC-F14; Q93TU3, primase-like protein, broad-host-range IncQ-like plasmid pTC-F14. (*B*) Nucleotide sequence alignment of *ssiA* (*Upper*) and *ssiB* (*Lower*) DNA segments from RSF1010 with *ssiA* and *ssiB*-like DNA segments found in other plasmids. The nucleotide sequence search was accomplished with blastN (34) and aligned with CLUSTALW (31), and figures were prepared with ESPript (32). Identical nucleotides shared by the aligned sequences are in white letters on red background, and identical nucleotides not shared by all sequences are in red letters. + indicates coding strand, − indicates noncoding strand. The priming site in *ssiA* DNA is marked by *. The numbers behind the plasmid names indicate the positions on the coding or noncoding strands. Sequence identity for *ssiA* and *ssiA*-like DNA segments is 50% and 90%, respectively, when pXF5823 was omitted from the alignment; sequence identity for *ssiB* was 60% and 90% when pXF5823 was omitted from the alignment. Plasmids pRAS3, pTF-FC2, pRSB105, and pTC-F14 (bottom group in *A*) encoding proteins related to RepB′ do not share DNA-segments homologous to *ssiA* and *ssiB*, but might have equivalent palindromic priming sites.

Table S1. Data collection and refinement statistics, RepB′ structure

Statistic	Native crystal	OsO_3-derivative crystal			
		Peak	Inflection	High remote	Low remote
Data collection					
Wavelength (Å)	0.931	1.1402	1.1406	1.0442	1.2557
Space group	P43212		P43212		
Unit cell (Å)					
a = b	91.74		90.4		
c	83.48		83.2		
$\alpha = \beta = \gamma$ (°)	90		90		
Resolution (Å)	50–1.98		50–2.7		
$I/\sigma(I)$	19.5 (3.4)	30.94 (2.3)	33 (3.7)	37.17 (4.0)	30.24 (2.1)
Completeness (%)	99.8 (98.4)	96.5 (72.5)	99.2 (99.9)	99.9 (100)	96 (64.4)
R_{sym} (%)*	7.6 (53.3)	3.9 (27.4)	3.9 (34.9)	3.7 (33.9)	4 (36.1)
Reflections measured	209,883	62,854	70,362	70,980	61,428
Unique reflections	24,819	17,528	18,120	18,159	17,497
Multiplicity	8.5	3.6	3.9	3.9	3.5
Refinement					
Resolution (Å)	19.2–1.98				
Reflections used for reference	24,774				
Working set size (%)	23,748 (95.8)				
Test set size (%)	1026 (4.1)				
R_{work}/R_{free}	19.5/23.3				
No. atoms					
Protein	2,244				
Water oxygen	253				
Sulfate (2)	10				
Pyrophosphate (1)	9				
Overall mean *B* value (Å^2)	36.1				
rms					
Bond lengths (Å)	0.010				
Bond angles (°)	1.065				
Ramachandran plot					
Most favored (%)	95.2				
Additionally allowed (%)	4.4				
Generously allowed (%)	0.4				

Highest-resolution shell is shown in parentheses.

*$R_{sym} = \Sigma \mid I - \langle I \rangle \mid / \Sigma \langle I \rangle$, where I is the observed and $\langle I \rangle$ is the average intensity of the given reflection.

Table S2. Conserved amino acids within the catalytic center of RepB′

RepB′	LigD Pol	Pfu-primase	Pho-primase	Sso-primase
Asp77	Asp669	Asp95	Asp95	Asp101
Asp78	Asp671	Asp97	Asp97	Asp103
Asp134	Asp759	Asp280	Asp280	Asp235
Arg145	Arg776	Arg292	Arg292	Arg247
Tyr62	His651	Tyr72	Tyr72	Phe74
Leu75	Val668	Val93	Val93	Leu99
Thr100	Ser704	Ser146	Ser146	Ser173
Gln106	His710	His151	His151	His179
His142	Arg762	Arg289	Arg289	Arg247
Asn151	Arg778	His298	His298	His253

Identical and type conserved (italic) amino acids of RepB′ within the compared structures are shown.

Table S3. Dissociation constants of complexes formed by RepB′ catalytic and helix-bundle domains to *ssiA* and to *ssiA*(3′Δ13) DNA

K_d	RepB′	Catalytic domain	Helix-bundle domain
ssiA	3.59 (±2.0) μM	2.01 (± 0.4) μM	27.0 (± 1.2) μM
ssiA(3′Δ13)	7.95 (± 0.98) μM	5.22 (±0.98) μM	25.8 (± 2.7) μM

Dissociation constants of complexes formed by RepB′, N-terminal catalytic domain (residues 1–212), or helix-bundle domain (residues 212–323) to intact *ssiA* DNA or *ssiA*(3′Δ13) were determined by analytical ultracentrifugation. rmsd are denoted in parentheses. The measurements show that RepB′ binds to intact *ssiA* DNA as monomer.

Table S4. Data collection and refinement statistics, catalytic domain/*ssiA*(3′Δ13) DNA structure

Data collection	
Wavelength (Å)	0.9184
Space group	$P4_32_12$
Unit cell (Å)	
a = b	85.3
c	68.8
$\alpha = \beta = \gamma$ (°)	90
Resolution (Å)	19.0–2.7
$I/\sigma(I)$	16.5 (5.6)
Completeness (%)	97.5(95)
R_{sym} (%)*	11.1 (39.4)
Reflections measured	52,296
Unique reflections	7,215
Refinement	
Resolution (Å)	19.0–2.7
Reflections used for refinement	6,997
Working set size (%)	95.0
Test set size (%)	5.0
R_{work}/R_{free}	21.4/27.0
No. atoms	
Protein + DNA	1,986
Water oxygen	24
Overall mean B value (Å^2)	31.7
rmsd	
Bond lengths (Å)	0.005
Bond angles (°)	1.0
Most favored Ramachandran plot (%)	95.1
Additionally allowed (%)	4.3
Generously allowed (%)	0.6

Highest-resolution shell is shown in parentheses.

*$R_{sym} = \Sigma \mid I - \langle I \rangle \mid / \Sigma \langle I \rangle$, where I is the observed, and $\langle I \rangle$ is the average intensity of the symmetry equivalent reflections.