

Crystal structure of a prokaryotic replication initiator protein bound to DNA at 2.6 Å resolution

Hirofumi Komori, Fujihiko Matsunaga¹,
Yoshiki Higuchi, Masamichi Ishiai^{1,2},
Chieko Wada¹ and Kunio Miki³

Department of Chemistry, Graduate School of Science, Kyoto University, Sakyo-ku, Kyoto 606-8502 and ¹Institute for Virus Research, Kyoto University, Sakyo-ku, Kyoto 606-8507, Japan

²Present address: Institute for Hepatic Research, Kansai Medical University, Osaka 570, Japan

³Corresponding author
e-mail: miki@kuchem.kyoto-u.ac.jp

The initiator protein (RepE) of F factor, a plasmid involved in sexual conjugation in *Escherichia coli*, has dual functions during the initiation of DNA replication which are determined by whether it exists as a dimer or as a monomer. A RepE monomer functions as a replication initiator, but a RepE dimer functions as an autogenous repressor. We have solved the crystal structure of the RepE monomer bound to an iteron DNA sequence of the replication origin of plasmid F. The RepE monomer consists of topologically similar N- and C-terminal domains related to each other by internal pseudo 2-fold symmetry, despite the lack of amino acid similarities between the domains. Both domains bind to the two major grooves of the iteron (19 bp) with different binding affinities. The C-terminal domain plays the leading role in this binding, while the N-terminal domain has an additional role in RepE dimerization. The structure also suggests that superhelical DNA induced at the origin of plasmid F by four RepEs and one HU dimer has an essential role in the initiation of DNA replication.

Keywords: autogenous repressor/crystal structure/
DNA-binding protein/RepE–iteron complex/replication
initiator

Introduction

The DNA replication of bacteria, bacteriophages and many plasmids is initiated by the binding of initiator proteins to specific binding sites at replicative origins. This binding promotes the localized unwinding of the origin DNA. Afterwards, helicase is directed to the single stranded region and a prepriming complex for priming and DNA synthesis is generated (Bramhill and Kornberg, 1988). Initiation of DNA replication by initiator binding to origin sequences is an important element of ordered cell proliferation. However, little is known about the structural basis of the mechanism of initiation of DNA replication because most initiator proteins generally tend to aggregate easily, even at low concentrations, making the growth of crystals for structural analysis very difficult.

Mini-F plasmid, a small derivative of the F plasmid of *Escherichia coli*, provides a simple model for the analysis of the mechanism of initiation of DNA replication. Mini-F plasmid is maintained at 1–2 copies per host chromosome and its replication is stringently controlled at the level of initiation. Mini-F contains a set of genes required for its characteristic mode of replication and partition including an origin of replication (*ori2*), the *repE* gene encoding the initiator protein, the incompatibility gene *incC* and the partition genes *sop* (*par*) (Kline, 1985). In addition to *ori2* and RepE, the host factors DnaA and HU, which are essential to the initiation of chromosomal DNA replication (*oriC*), are required for the initiation of replication of mini-F (Hansen and Yarmolinsky, 1986; Kline *et al.*, 1986; Murakami *et al.*, 1987; Wada *et al.*, 1988; Ogura *et al.*, 1990). The *ori2* (*incB*) region contains four directly repeated sequences of 19 bp, an AT-rich region and binding sites for DnaA protein (Figure 1). Similar sets of DNA sequences have been found near the replication origins of several other plasmids and the bacterial chromosome, and appear to play fundamental roles in the initiation of DNA replication.

The regulatory mechanism of replication of mini-F plasmid has been well studied (Ishiai *et al.*, 1994). The RepE protein (251 residues, 29 kDa) plays an essential role in the initiation of mini-F plasmid replication. The monomeric and dimeric forms of RepE have distinctive functions, i.e. initiation of replication and autogenous repression, respectively (Ishiai *et al.*, 1994). RepE monomers bind to the four direct repeats (iterons) within the origin (*ori2*) region and RepE dimers bind to the inverted repeats of the *repE* operator. There are common 8 bp sequences in both the iterons and the *repE* operator (Figure 1). RepE exists mostly as a dimer within the cell, and conversion of this stable inactive form into the active monomeric form for initiation requires the action of host cell chaperones DnaK, DnaJ and GrpE (Kawasaki *et al.*, 1990, 1991; C.Wada, unpublished data; see Figure 1).

Previous genetic and biochemical studies of the plasmid initiator protein RepE have afforded some information about its structural features (Matsunaga *et al.*, 1995, 1997). A helix–turn–helix (HTH) motif, known as the DNA binding motif (Masson and Ray, 1986; Brennan and Matthews, 1989; Matsunaga *et al.*, 1995), has been assigned by computer analysis to the internal region of RepE (residues 64–83) (Figure 2), whereas the critical region for binding both to *ori2* and to the operator is located in the C-terminal region (residues 168–242) (Matsunaga *et al.*, 1995). No other DNA binding motifs have been found in the C-terminal region of RepE. A leucine-zipper (LZ) motif generally implicated in protein–protein interactions is expected to exist in the N-terminal regions of several plasmid initiator proteins (residues 21–55 of RepE) (Giraldo *et al.*, 1989; Matsunaga *et al.*,

1995). The LZ motif of the initiator protein of pPS10 plasmid, RepA, which has sequence similarity with RepE, is involved in its dimerization (Garcia *et al.*, 1996). In mini-F, a number of mutations that affect dimerization of RepE are located within the central region (residues 93–135; Matsunaga *et al.*, 1997). The actual function of the LZ motif of RepE has remained obscure.

We determined the crystal structure of the RepE monomer complexed to an iteron DNA sequence, in order to establish the structural basis for the dual functions of RepE and to understand the regulatory mechanism of mini-F replication. This paper presents the first three-dimensional structure of a prokaryotic initiator protein.

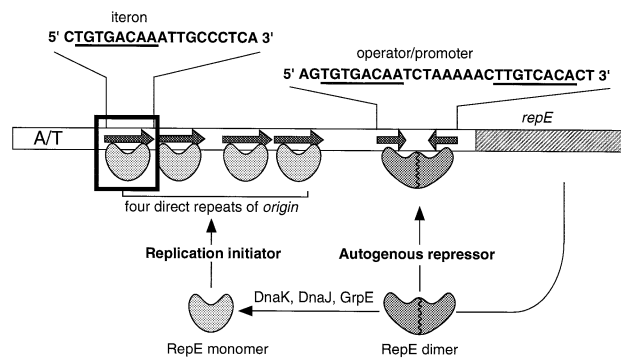


Fig. 1. A schematic drawing of the functions of the RepE initiator protein in mini-F plasmid replication. The RepE monomers bind to the four iterons (direct repeats) of *ori2* to initiate replication, whereas the RepE dimers bind to the inverted repeat of the *repE* promoter–operator to repress *repE* transcription. Parts of the repeated sequences (iterons) are shown at the top of the Figure where portions shared by the direct and inverted repeats are underlined (common 8 bp). The box indicates the RepE54–iteron DNA complex determined in this study.

Results and discussion

Structure determination

A RepE mutant, RepE54, which carries a point mutation in its central region (Arg118→Pro replacement), has markedly enhanced initiator activity but little or no repressor activity. The mutant protein is stable in the monomeric form without chaperones, fails to aggregate even at a high concentration (~40 mg/ml) and binds to iterons with great efficiency. In contrast, the wild-type RepE (monomer and dimer) protein tends to aggregate easily (Wada *et al.*, unpublished data). Therefore, the RepE54 appeared suitable, not only for the preparation of the RepE–iteron complex, but also for the preparation of RepE54 monomer complexed to DNA (Figure 3A) (Komori *et al.*, 1999). The crystal structure was solved by the multiple isomorphous replacement method and refined to $R = 0.213$ ($R_{\text{free}} = 0.274$) at 2.6 Å resolution (Table I). The present model comprises residues 15–246 of RepE54 and 21 bp of the iteron DNA together with 20 water molecules and a Mg^{2+} ion (Figure 3B and C). The overhanging thymines (T22 and T44) of the iteron DNA, and a few residues at the N- and C-termini and in the loop regions (residues 50–55 and 98–109) of RepE54, are not included in the present model due to their disordered structure.

Overall structure of the RepE54 protein

The crystal structure of the RepE54–iteron complex (Figure 3B and C) showed a novel type of protein–DNA binding. RepE54 is composed of two distinct N- and C-terminal domains which are structurally similar and related to each other by a non-crystallographic dyad, although no such similarity was expected from its amino acid sequence.

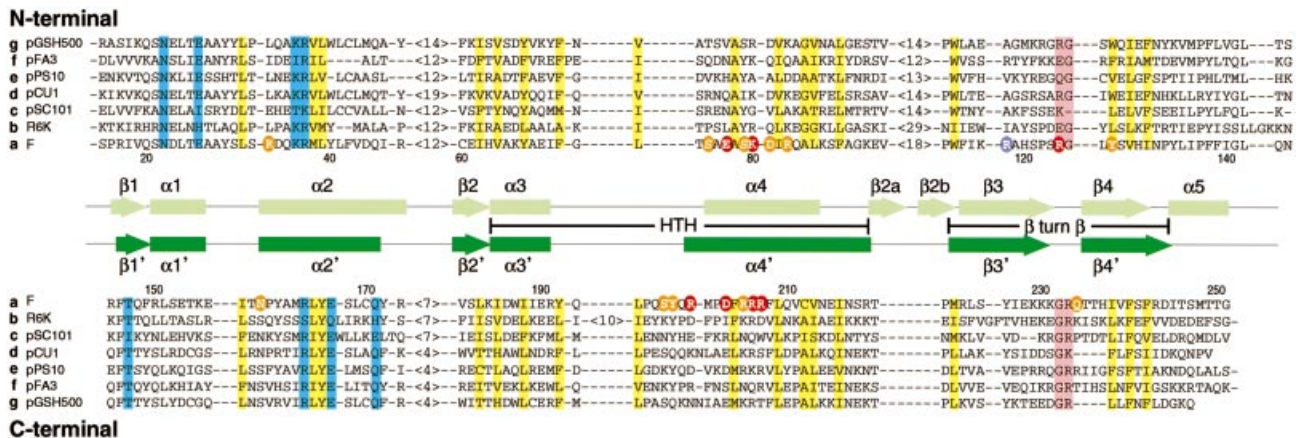


Fig. 2. Comparison of the amino acid sequences of the N- and C-terminal regions of plasmid initiator proteins on the basis of the three-dimensional structure of RepE. **a**, RepE (*E. coli*); **b**, π (R6K, *E. coli*); **c**, RepA (pSC101, *E. coli*); **d**, ORF (pCU1, *E. coli*); **e**, RepA (pPS10, *Pseudomonas syringae*); **f**, 39K basic protein (pFA3, *Neisseria gonorrhoeae*); **g**, RepB (pGSH500, *Klebsiella pneumoniae*) (Murotsu *et al.*, 1981; Armstrong *et al.*, 1984; Germino and Bastia, 1984; Gilbride and Brunton, 1990; Krishnan *et al.*, 1990; Nieto *et al.*, 1992; da Silva-Tatley and Steyn, 1993). Hyphens indicate gaps. The terminal sequences of some of the proteins are not shown. The numbers in < > indicate the number of residues that are not displayed. The secondary structure elements in the present crystal structure of RepE54 (118Arg→Pro) were assigned. The α -helical segments are shown as cylinders and the β -strands as arrows. The N-terminal domain (residues 15–144) and the C-terminal domain (residues 145–246) are shown in light green and dark green, respectively. The residues in contact with bases and the phosphate backbone of the iteron DNA are shown in red and orange, respectively. The mutation point of RepE54 (118R→P) is shown in violet. The hydrophobic conserved residues are highlighted in yellow. The conserved Arg–Gly sequence in the β –turn– β motifs of both domains are in pink. The polar amino acid residues responsible for interactions between N- and C-terminal domains are in blue. Asn22, Glu26 and Lys36 in the N-terminal domain interact with Thr147, Glu167 and Gln171 in the C-terminal domain, respectively.

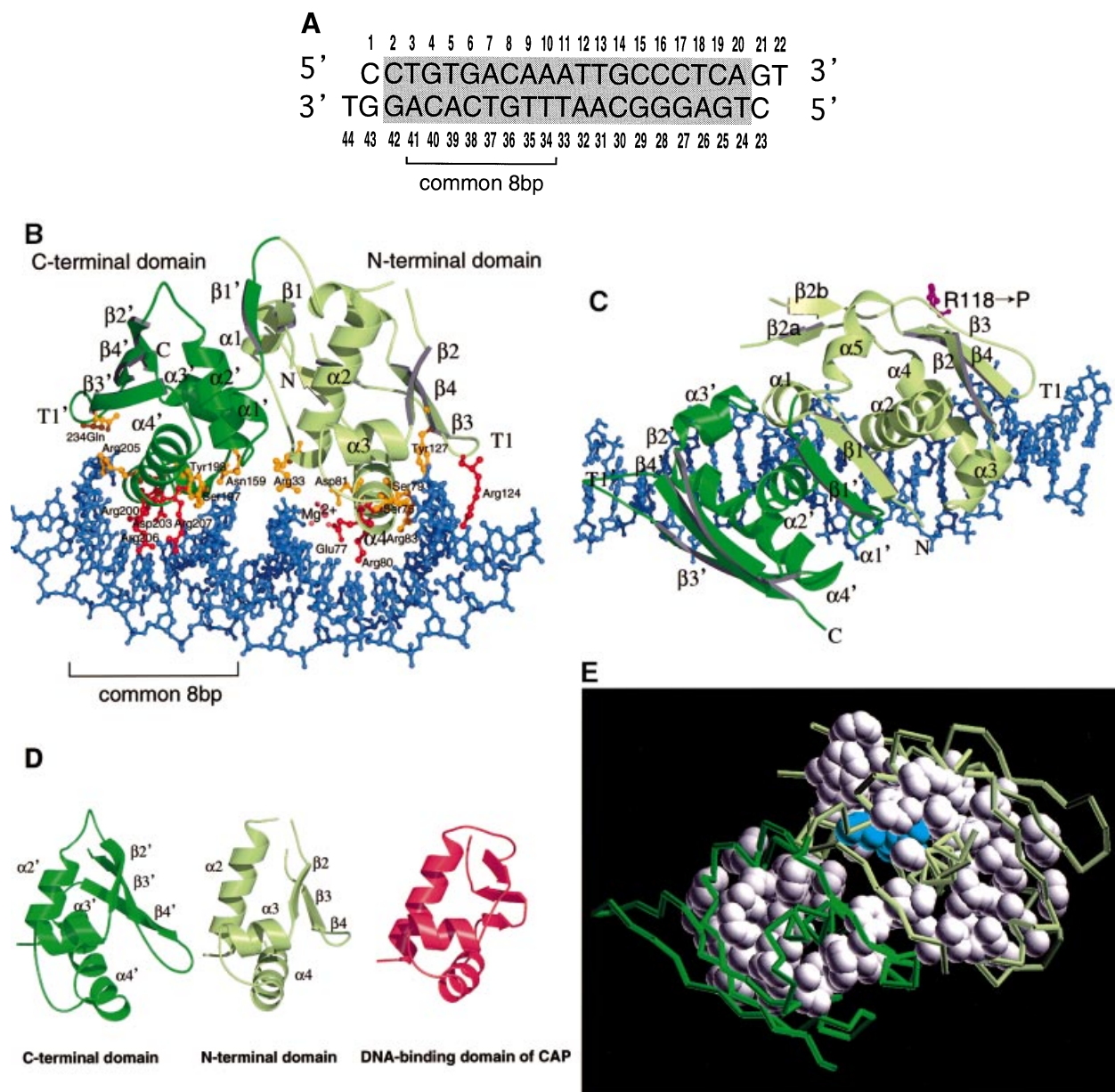


Fig. 3. (A) A synthetic 21 bp DNA duplex with 3' overhanging thymines used for co-crystallization with RepE. The shaded box indicates the 19 bp iteron sequence. The 8 bp TGTGACAA (3–10) sequence appears in both iterons and operators. (B) The overall structure of the RepE54–iteron complex. The N- and C-termini of RepE54 are labeled N and C, respectively. The coloring scheme is the same as that in Figure 2. (C) A view rotated 90° around the horizontal axis relative to (B). This Figure was generated by MOLSCRIPT (Kraulis, 1991) and Raster3D (Merritt and Murphy, 1994). (D) Comparison of both N- and C-terminal domains of RepE54 with the DNA-binding domain of CAP. The N-terminal domain ($\alpha 2$ – $\alpha 4$ and $\beta 2$ – $\beta 4$) is shown in light green, the C-terminal domain ($\alpha 2'$ – $\alpha 4'$ and $\beta 2'$ – $\beta 4'$) in dark green and the DNA-binding domain of CAP in red. (E) A view of the hydrophobic core in both N- and C-terminal domains. The hydrophobic residues of RepE54, which form an internal hydrophobic core, are shown as a white space-filling presentation. The conserved leucines in the N-terminus buried within the hydrophobic core are shown in cyan.

Both domains are comprised of four α -helices ($\alpha 1$ – $\alpha 4$ for the N-terminal domain and $\alpha 1'$ – $\alpha 4'$ for the C-terminal domain) and four β -strands ($\beta 1$ – $\beta 4$ and $\beta 1'$ – $\beta 4'$). The $\beta 3$ strand in the N-terminal domain is somewhat distorted where the mutation point of RepE54 (Arg118→Pro) is located. Additional secondary structures ($\beta 2a$, $\beta 2b$ and $\alpha 5$) have been assigned to the N-terminal domain (Figures 2 and 3C). As expected for a prokaryotic DNA binding motif, there is a canonical HTH motif in the N-terminal domain containing a four-residue turn with glycine at the second position ($\alpha 3$ –FGLT– $\alpha 4$). In the corresponding region of the C-terminal domain, which is related to the N-terminal domain by a pseudo dyad, a similar HTH

motif can be found but the four-residue turn is replaced by eight residues ($\alpha 3'$ –YQLPQSYQ– $\alpha 4'$).

Each domain of RepE54 displays topologically similar folding and contains a winged-helix motif like the one found in the DNA-binding domain of the catabolite gene activator protein (CAP) (Schultz *et al.*, 1991; Brennan, 1993) (Figure 3D). The domain is composed of three α -helices ($\alpha 2$ – $\alpha 4$ and $\alpha 2'$ – $\alpha 4'$) flanked by three β -strands ($\beta 2$ – $\beta 4$ and $\beta 2'$ – $\beta 4'$). In particular, the folding of the N-terminal domain is structurally similar to CAP, the C_{α} r.m.s. being 2.88 Å (2.15 Å for the three α -helices and 0.84 Å for the HTH motif).

In Figure 3E, each domain of RepE54 is shown to have

Table I. Crystallographic data statistics

Data collection Data set	Source ^a	Wavelength (Å)	Resolution (Å)	Reflections			$\langle I/\sigma(I) \rangle$	R_{merge} (%) ^b
				Total observed	Unique	Completeness (%)		
Nati1	BL18B	1.000	35–2.6	57565	15208	88.7	34.7	3.5
Nati2	BL41XU	0.708	35–2.6	46882	15420	89.1	22.7	4.1
Hg1	BL18B	1.000	35–3.0	23555	9199	72.4	14.8	6.3
Hg2	Raxis	1.542	35–3.0	22111	9803	79.5	13.0	7.2
IdU12	Raxis	1.542	35–3.1	28648	9856	86.7	11.8	8.3
IdU13	Raxis	1.542	35–3.0	29418	10501	89.4	16.8	6.2
IdU33	BL6A	1.000	35–3.0	30921	9481	80.4	20.5	5.5

Phasing statistics						
Derivative	R_{iso} (%) ^c	No. of sites	Phasing power ^d		Cullis $R_{\text{anomalous}}$ ^f	Mean Figure of merit
			Acentric/Centric	Acentric/Centric		Acentric/Centric
Hg1	17.0	2	2.28/1.60	0.58/0.54	0.83	
Hg2	16.3	2	1.93/1.45	0.63/0.61	0.95	
IdU12	15.8	1	0.99/1.13	0.82/0.66	0.96	
IdU13	12.5	1	1.49/1.43	0.69/0.58	0.92	
IdU33	10.6	1	1.23/1.30	0.75/0.59	0.98	
						0.634/0.790

Refinement statistics				
Resolution (Å)	No. of reflections	R (%) ^g	R_{free} (%) ^g	
6.0–2.6	13738	21.3	27.4	

	No. of non-hydrogen atoms	r.m.s. deviation		B_{ave} (Å ²)
		Bonds (Å)	Angles (°)	
Protein	1783	0.007	1.170	30.6
DNA	855	0.005	1.199	31.8
Solvent molecules	21			20.1

^aBL18B, BL6A: Photon Factory, BL41XU: Spring-8, Raxis: Rigaku R-AXIS IV with an ultra18X generator at our laboratory.

^b $R_{\text{merge}} = \sum |I_i - \langle I_i \rangle| / \sum \langle I_i \rangle$, where I_i is the observed intensity and $\langle I_i \rangle$ is the average intensity over symmetry equivalent measurements.

^c $R_{\text{iso}} = \sum ||F_{\text{PH}}| - |F_{\text{P}}|| / \sum |F_{\text{P}}|$, where F_{P} and F_{PH} are the derivative and native structure factors, respectively.

^dPhasing power = $\sum |F_{\text{H}}| / \sum ||F_{\text{PHobs}}| - |F_{\text{PHcalc}}||$, where F_{PH} and F_{H} are the derivative and calculated heavy-atom structure factors, respectively.

^eCullis $R = \sum ||F_{\text{PH}\pm F_{\text{P}}}| - |F_{\text{H}}|| / \sum |F_{\text{PH}\pm F_{\text{P}}}|$, where F_{PH} , F_{P} and F_{H} are the derivative, native and calculated heavy-atom structure factors, respectively.

^fCullis $R_{\text{anomalous}} = \sum |F_{\text{PH}(+)} - F_{\text{PH}(-)}| - 2F_{\text{H}} \sin \alpha_{\text{p}} / \sum |F_{\text{PH}(+)} + F_{\text{PH}(-)}|$, where α_{p} is the calculated protein phase.

^g $R = \sum ||F_{\text{obs}}| - |F_{\text{calc}}|| / \sum |F_{\text{obs}}|$. R_{free} is the same as R , but for a 5% subset of all reflections that were never used in crystallographic refinement. r.m.s., root mean square.

an internal hydrophobic core. The leucine residues, which were expected to form an LZ motif, are present within the hydrophobic core in the N-terminal domain, but do not constitute a canonical LZ motif. In the RepE monomer structure, the leucine residues are not exposed on the surface but are responsible for maintaining the tertiary structure in a stable configuration.

Iteron DNA conformation

The iteron DNA adopts a B-form-like conformation (Figure 3B and C) and stacks in a head-to-tail manner in the crystal. The average twist and rise parameters for two successive base pairs are 34.1° and 3.4 Å, respectively, which corresponds to ~36 Å per turn of DNA. The largest roll angle is 13.0°. Overall, the iteron DNA is slightly bent (~20°) in the present crystal structure. The gel shift assay study also suggests such bending for RepE bound to a single iteron, where a bending angle of ~50° was estimated (Kawasaki *et al.*, 1996). This comparatively large value is probably due to the use of much longer

DNA fragments carrying an iteron (140 bp) than those used in the crystals.

Protein–DNA interactions

Figure 4A summarizes all the RepE54–iteron contacts. There are extensive polar interactions which exist mainly between the two recognition helices of both domains (α_4 , residues 75–87, and α_4' , residues 200–219 in Figure 2) and the major groove of the iteron DNA. α_4 is the second helix of the HTH motif in the N-terminal domain. Only two residues (Glu77 and Lys80) in the α_4 helix interact directly with bases on the DNA (C15 and G29; Figure 4B). The other polar residues point towards the major groove but they are not close enough for direct interaction with bases on the DNA. Lys80, which has a high B-factor, binds poorly to G29. On the other hand, α_4' is the second helix of the non-canonical HTH motif in the C-terminal domain (Figure 4C). This α -helix lies in the major groove containing the 8 bp sequence which is common to both the iteron and the operator of the *repE* gene (Figure 3A).

The $\alpha 4'$ helix has a kink at its center (residue 207) which would promote close fitting of this helix into the major groove containing the common 8 bp sequence. The unique form of the HTH motif enables the recognition helix, $\alpha 4'$, to engender multiple specific contacts (Arg200, Asp203, Arg206 and Arg207) with the bases on the DNA (G4, T5, C38, G6 and G36). We found that iterons carrying mutations at G4, T5 or G6 did not bind to RepE *in vitro* (H.Uga and C.Wada, unpublished data). Since the 8 bp sequence (TGTGACAA) in the iteron is also found in the operator DNA sequence, specific DNA binding by the C-terminal domain could contribute to RepE interactions with both the iterons and the operator DNA. This is consistent with mutation analysis (Matsunaga *et al.*, 1995). Many mutations (point and double mutations) that severely affect both *ori2*- and operator-binding activities have been identified within the C-terminal domain, including the $\alpha 4'$ helix (168–242 residues; Figure 4D). Most of the mutations are present within the hydrophobic core of the C-terminal domain. These mutations might disrupt the structure of the C-terminal domain. This structural disruption of the C-terminal domain containing the $\alpha 4'$ helix is predicted to strongly affect specific DNA binding.

Both recognition helices also make contacts with the phosphate backbone of the iteron DNA, but the numbers of amino acids implicated in these contacts differ between the two domains; three (Ser75, Ser79 and Arg83) are involved in the N-terminal domain, but only one (Arg205) in the C-terminal domain. Some additional contacts of RepE54 with DNA are made by helices $\alpha 2$ and $\alpha 2'$, and turns T1 and T1'. Arg33 in the $\alpha 2$ helix of the N-terminal domain makes direct contact with the phosphate backbone of T13 at the center of the iteron DNA, whereas Asn159 of the $\alpha 2'$ helix interacts with the phosphate of T35 through direct and water-mediated contacts. Arg124 at the T1 turn in the N-terminal domain protrudes into the minor groove and makes contact with the base G25, whereas the corresponding T1' turn in the C-terminal domain interacts with the phosphate backbone. The DNA interactions of the strand–turn–strand motif, in addition to that of the HTH motif, are reminiscent of those found in winged-helix motif proteins such as CAP.

The Mg^{2+} ion was found in the cavity between the protein and the DNA near the $\alpha 4$ helix in the N-terminal domain. It is octahedrally coordinated by four water molecules and two residues (Glu77 and Asp81), and affords a few water-mediated DNA contacts (Figure 4A and B). The Mg^{2+} ion might indirectly assist the binding between the iteron and the $\alpha 4$ recognition helix which forms much looser protein–DNA contacts than the $\alpha 4'$ helix in the C-terminal domain.

Implications for dimerization

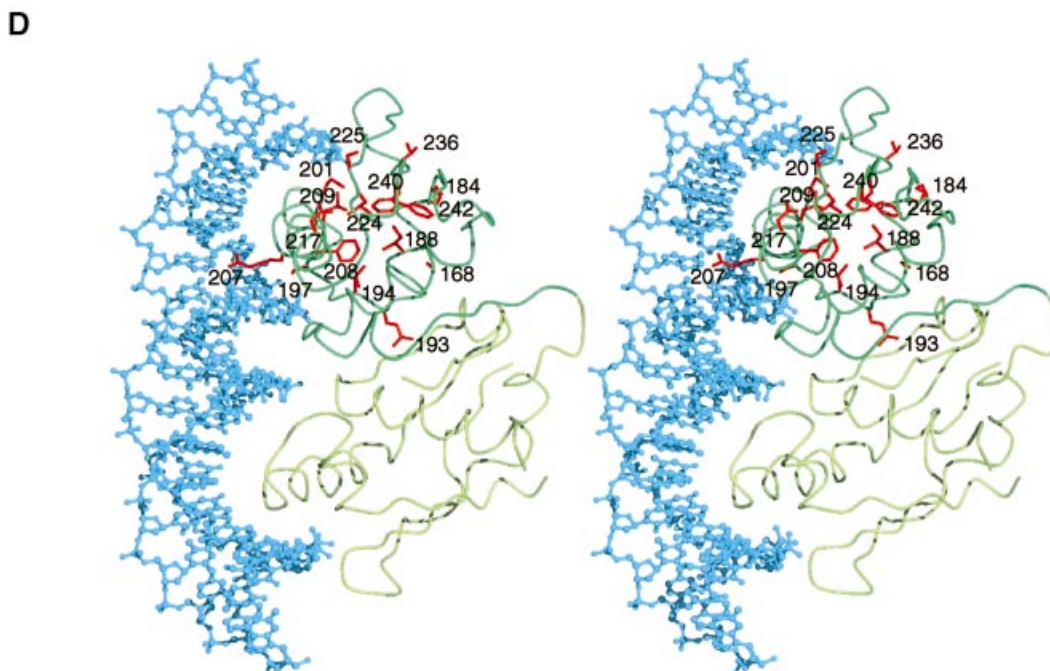
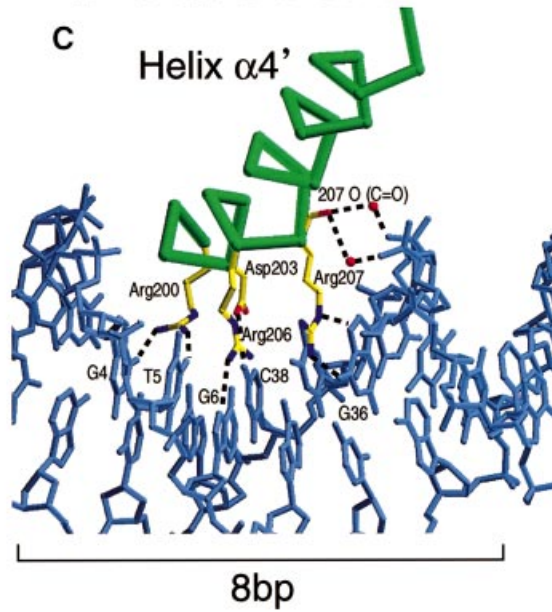
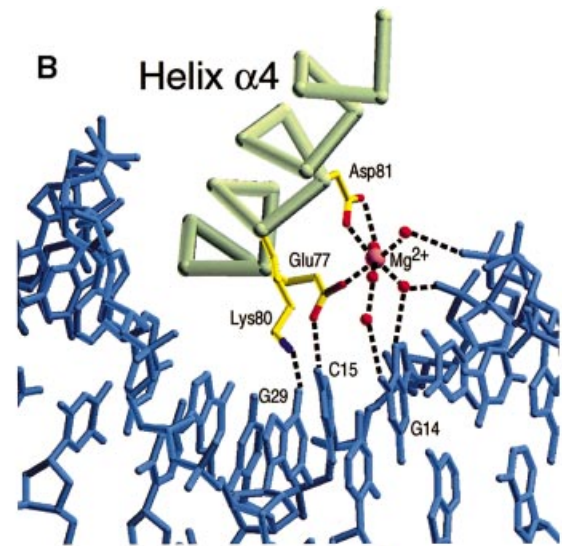
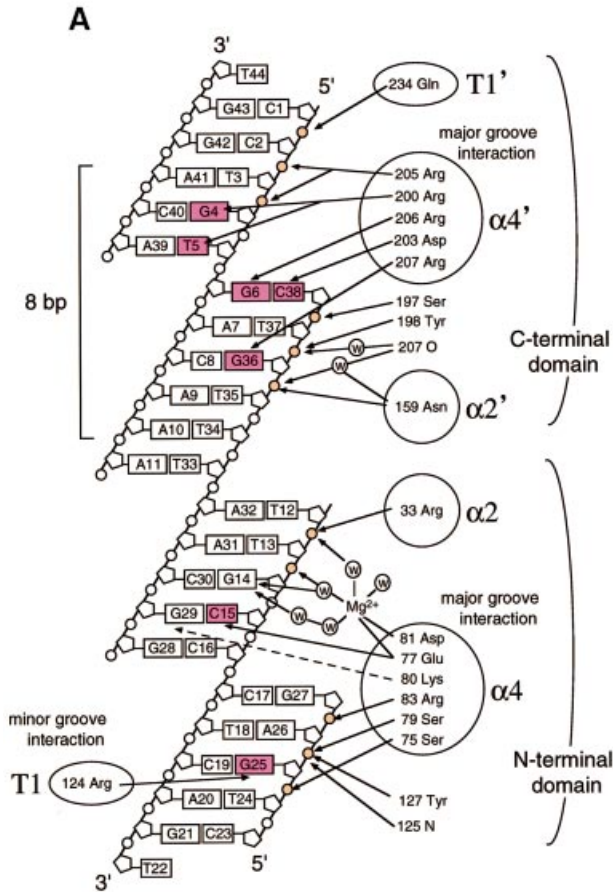
Wild-type RepE usually exists in dimeric form and recognizes the operator of *repE* where it acts as an autogenous repressor (Ishiai *et al.*, 1994). As mentioned above, the recognition helix $\alpha 4'$ of the HTH motif in the C-terminal domain is responsible for RepE binding to both the iteron and operator DNA sequences. Therefore, the $\alpha 4'$ helix in dimeric RepE would also be expected to bind to the operator DNA in a similar manner to that of the RepE54–iteron complex. It is clear from this viewpoint that the N-terminal domains of the dimer bound to the operator

will face each other and be responsible for dimerization. When we construct a model of the RepE dimer bound to operator DNA along the lines of the RepE54–iteron complex, a large steric hindrance occurs in the major part of the N-terminal domains and, especially, the $\alpha 4$ helices compete with each other for the same major groove (Figure 5A). This implies that a marked conformational change of the RepE protein is necessary to accommodate its dimeric form. Structural changes provided by the bending of the operator DNA do not appear to compensate for the steric hindrance between the RepE proteins. A hypothetical model for the RepE dimer would be possible if a large part of the N-terminal domain ($\beta 2$ – $\alpha 5$) were to flip out drastically. The loop connecting the C- and N-terminal domains and the disordered loop (residues 49–56) could be used as a hinge region (Figure 5B). In this model, the $\alpha 4$ helices of the HTH motifs are released from the major grooves and no longer interact with DNA. The $\alpha 4$ helices may be unnecessary for binding of the RepE dimer to the operator DNA, although they are responsible for binding the RepE monomer to the iteron DNA. In the RepE dimer, the two $\alpha 4'$ recognition helices would play more dominant roles in rigid interactions with the operator DNA.

The mutation at residue Pro118 of RepE54 is present within the $\beta 3$ strand. The other mutation sites found in dimerization defective RepE mutants (Matsunaga *et al.*, 1997) are also located between $\beta 2a$ and $\alpha 5$ (residues 93–135) including $\beta 3$ (Figure 5B). This region of the N-terminal domain may act as the dimer interface of RepE (Figure 5C). Furthermore, this region covers the conserved leucines at the N-terminus (residues 21–55) (Matsunaga *et al.*, 1995) (Figure 5B). Therefore, a drastic movement of this region and neighboring regions ($\beta 2$, $\beta 4$, $\alpha 3$ and $\alpha 4$) could expose the hydrophobic cluster of these leucines to the molecular surface, allowing them also to function as a dimer interface (Figure 5C). Consequently, the two hydrophobic cores of each N-terminal domain of the RepE monomers interact with each other in a face-to-face manner to form one hydrophobic core in the RepE dimer. The high stability of the RepE dimer may be partially caused by this additional hydrophobic interaction. This dimer model explains why the conserved leucines at the N-terminus have been implicated as residues of the dimerization interface in the initiator proteins of pPS10 plasmid (Garcia de Viedma *et al.*, 1996) although they are buried inside the RepE monomer structure (Figure 3D). During the conversion from dimer to monomer *in vivo*, there might be an intermediate state in which the hydrophobic core of the N-terminal domain is transiently exposed. It is known that chaperones bind to hydrophobic regions exposed on unstructured proteins to prevent their aggregation (Bukau and Horwich, 1998). It is also to be expected that chaperones stabilize such a hydrophobic intermediate state to facilitate conversion to the monomeric state.

Common dual structure for plasmid initiator proteins

The three-dimensional structures of other plasmid initiator proteins are not yet known; however, they display some similarities to RepE in their amino acid sequences (Matsunaga *et al.*, 1995). Some plasmid initiator proteins, such as those of pPS10 and pSC101, also have dual



functions as initiators and autogenous repressors like RepE (Matsunaga *et al.*, 1995; del Solar *et al.*, 1998). In this work, it was found that the RepE monomer has an internal dual structure and that many of the amino acid residues are related by 2-fold symmetry. In order to elucidate an

eventual structural relationship among plasmid initiator proteins, the N-terminal and C-terminal amino acid sequences of several plasmid initiator proteins were aligned together on the basis of the 2-fold symmetry of the three-dimensional structure of the RepE monomer

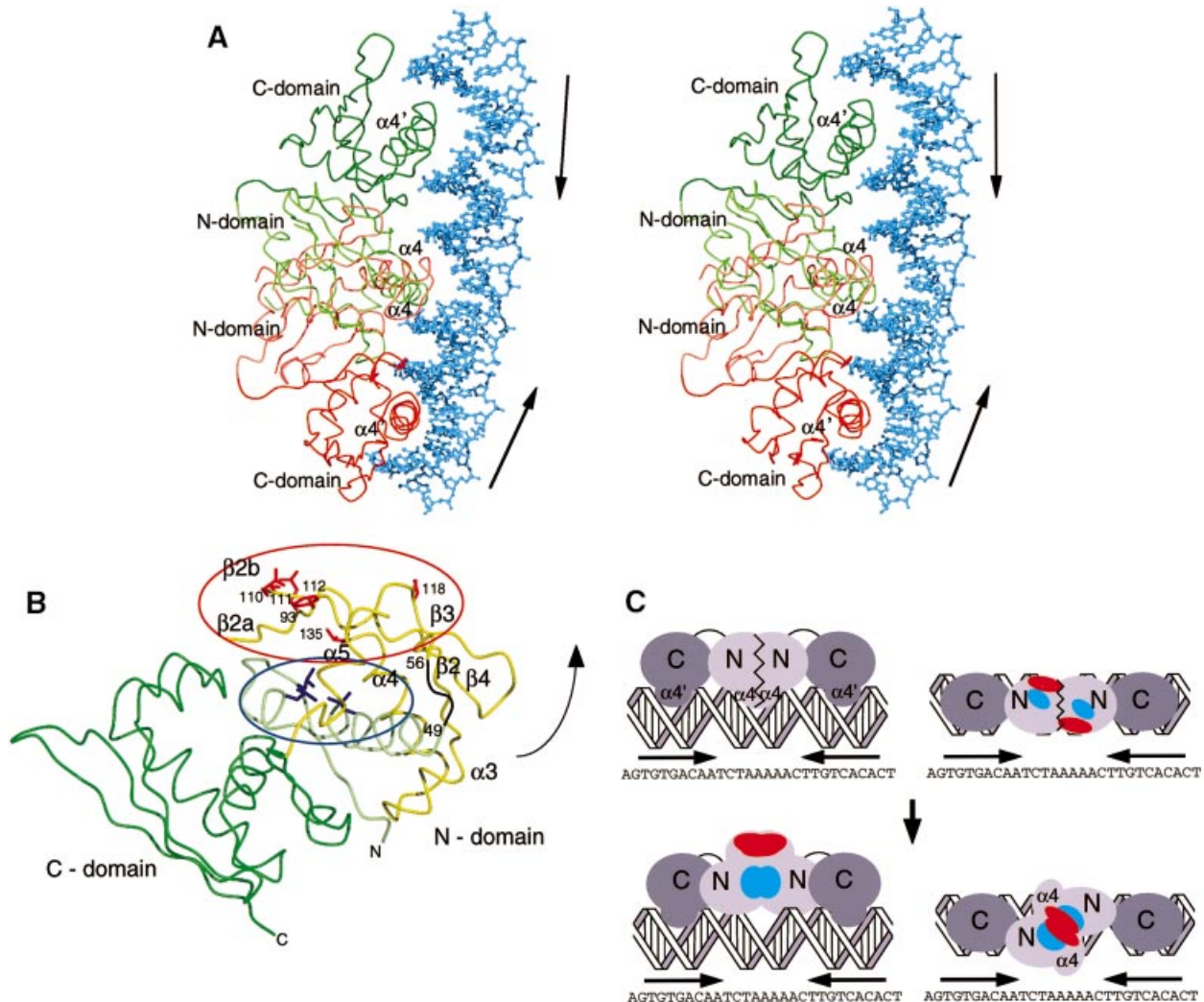


Fig. 5. RepE dimeric model. (A) Stereo view of a dimeric model of RepE without conformational changes. Two monomer structures are arranged in series on the operator DNA with the fixed protein conformation and protein–DNA interaction characteristics of the RepE54 (monomer)–iteron DNA complex. The two RepE monomers are shown in green and red. Two inverted arrows indicate the regions of the common 8 bp sequence. (B) Mutation positions of dimerization defective mutants. Mutation sites are shown in red and enclosed by a red circle. The conserved leucines in the N-terminus are shown in blue and enclosed by a blue circle. The yellow C_{α} chain indicates the region which is expected to flip out in the direction of the arrow. (C) Proposed RepE dimeric model with conformational changes of the N-terminal domains of the RepE monomer. Top: a schematic representation of a dimeric model without conformational change which is the same as that shown in (A). The regions of the common 8 bp sequence are also shown as two inverted arrows. N and C indicate the N- and C-terminal domains, respectively. The red and blue circles in the N-terminal domains indicate the mutation and conserved leucine regions shown in (B), respectively. Bottom: a dimeric model after conformational change of the N-terminal domains. In this model, both the mutation and conserved leucine regions can act as the dimer interface of RepE. In both top and bottom, the viewpoint of the right Figure has been rotated 90° along its horizontal axis with respect to that of the left figure.

Fig. 4. (A) A schematic review of RepE54–iteron interactions. The nucleotide numbering scheme is the same as in Figure 3A. The RepE54 and iteron interactions are indicated by arrows. Bases and phosphates in contact with RepE54 are shown in red and orange, respectively. Circled ‘W’s indicate water molecules. (B) and (C) Comparison of the protein–DNA interaction at helix $\alpha 4$ (light green) of the N-terminal domain and helix $\alpha 4'$ (dark green) of the C-terminal domain. Amino acid residues interacting with the DNA are shown in yellow, water molecules by red circles and the Mg^{2+} ion by a pink circle. (D) Mutation positions of DNA-binding-defective RepE mutants. Point mutations: 168S→P, 184K→I, 188I→N, 194L→P, 197S→R, 209L→P, 224L→P, 240F→S and 242F→S. Double mutations: 193Q→L and 201M→T, 201M→T and 225S→P, 207R→P and 208F→L, and 217N→D and 236T→S. All the point mutation residues (except for K184I) and one of the double mutation residues are directed towards the hydrophobic core of the C-terminal domain. On the other hand, other residues of the double mutations (Q193L, S225P, R207P and T236S) are exposed on the molecular surface. The positions of these mutations indicate that Q193L, S225P and T236S might not affect DNA binding activity and that the other mutations disrupt the structure of the C-terminal domain. The R207P mutation is located at the center of the recognition helix $\alpha 4'$ and could disrupt its structure.

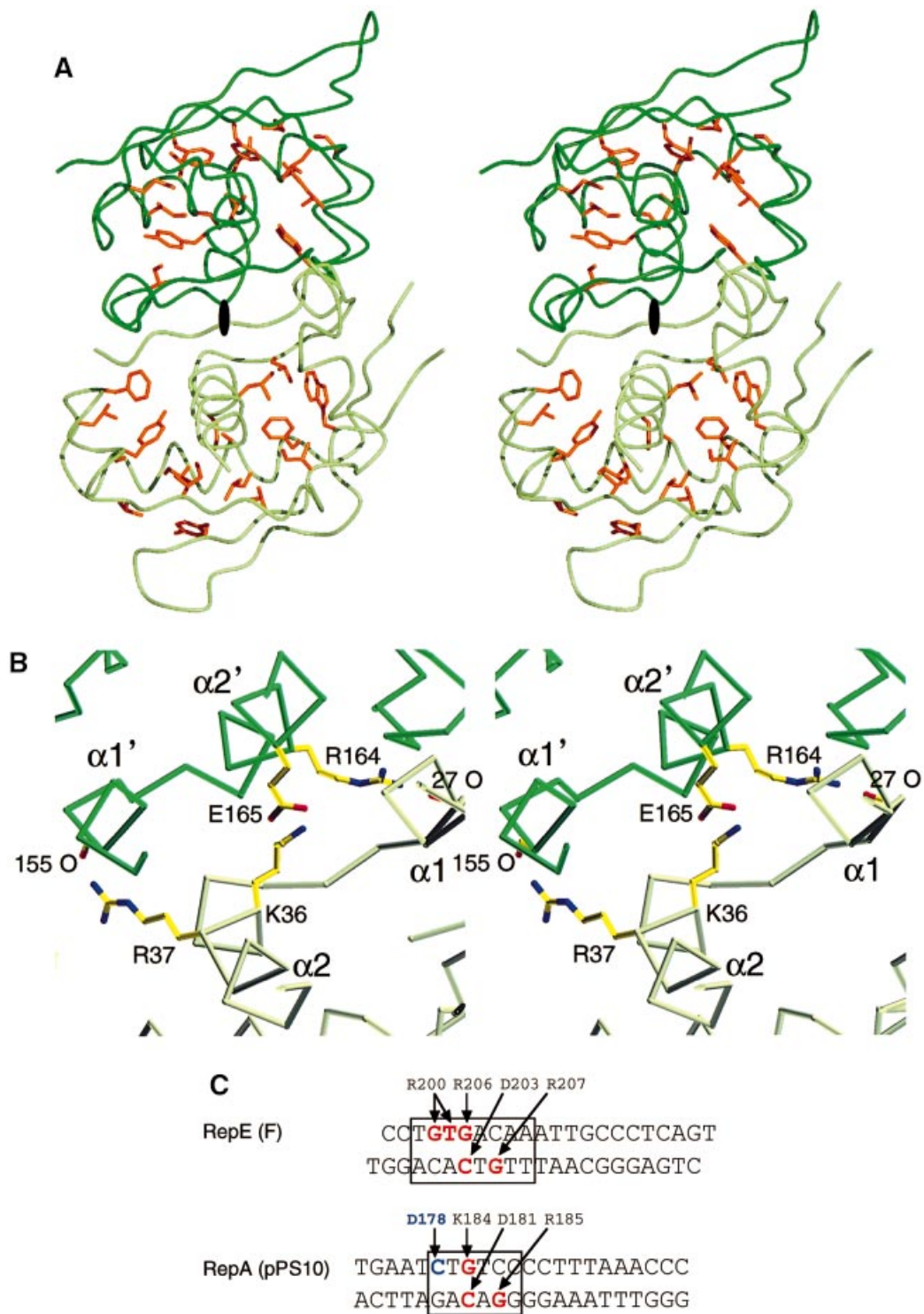


Fig. 6. (A) Positions of the conserved hydrophobic residues that form the hydrophobic core (in stereo). They are related by 2-fold symmetry, and the position of the 2-fold axis is indicated in black. (B) Polar interactions between the N- and C-terminal domains (in stereo). Arg37 in the $\alpha 2$ helix of the N-terminal domain interacts with the carbonyl oxygen of Lys155 in the $\alpha 1'$ helix of the C-terminal domain. Arg167 in the $\alpha 2'$ helix of the C-terminal domain interacts with the carbonyl oxygen of Ala27 in the $\alpha 1'$ helix of the N-terminal domain. (C) Comparison of the DNA binding site sequences (iteron) of RepE and RepA initiator protein of pPS10 plasmid. Conserved sequences with their operator DNA sequence are boxed. The corresponding amino acid residues of RepE and RepA in contact with the bases on the DNA are indicated.

(Figure 2). RepA, a familiar initiator protein of P1 plasmid, could not be aligned due to its poor sequence homology with RepE protein. Many hydrophobic residues related by 2-fold symmetry are found to be conserved in the other plasmid initiator proteins. All of these conserved hydrophobic residues are buried in the hydrophobic cores of

RepE (Figure 6A). This indicates that these hydrophobic residues in the other plasmid initiator proteins also have 2-fold symmetry and make two internal hydrophobic cores like RepE.

There are several internal polar interactions between the N- and C-terminal domains of the RepE monomer.

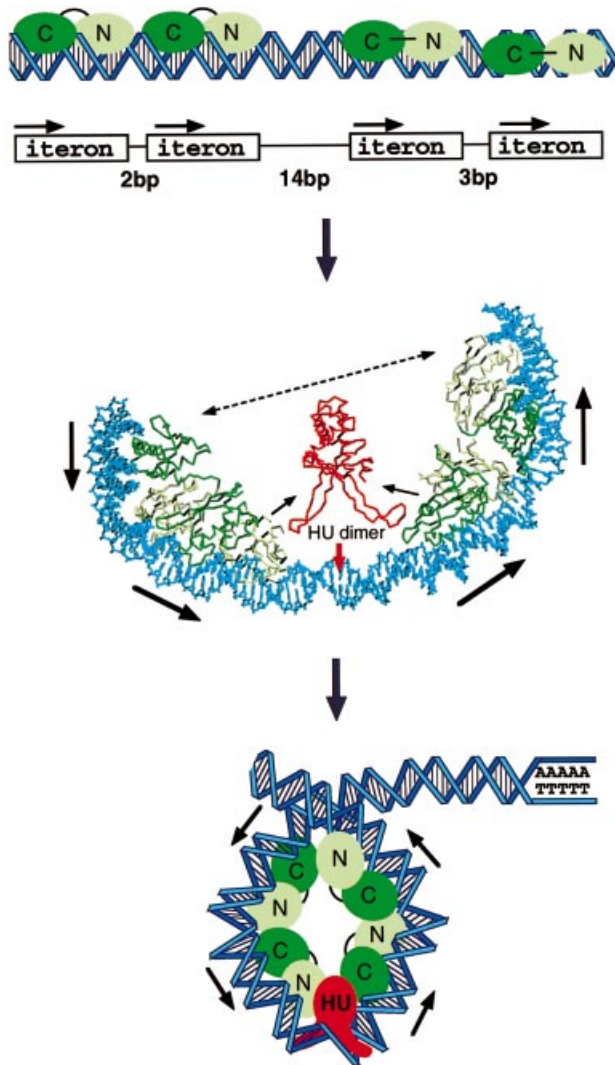


Fig. 7. Proposed superhelical DNA structure induced by the complex made up of four RepE monomers and one HU dimer. The N- and C-terminal domains of RepE54 are shown in light and dark green, respectively. Origin DNA is shown in blue. The regions of the common 8 bp sequence are also shown as arrows. The structure of the HU dimer (PDB: 1HUU; Tanaka *et al.*, 1984) is shown in red. The arrangement of the four iterons in the origin DNA shows the intervals between iterons (top). The four RepE monomers bind to four repeated iterons and each of two adjacent iterons can be bent about 90° to create contacts with two RepE monomers (middle). The HU dimer can make contact with the 14 bp region between the second and third iterons and interact with the two RepEs at the second and third iterons. As a result of these interactions, the origin DNA can be sharply bent at its 14 bp region and wound to form a superhelical structure, which could provoke unwinding of the neighboring AT-rich region for the initiation of DNA replication (bottom).

Arg37 and Arg164 in the $\alpha 2$ and $\alpha 2'$ helices, which are related by 2-fold symmetry, interact with $\alpha 1'$ and $\alpha 1$ helices, respectively (Figure 6B). These arginine residues are also maintained in the corresponding positions of the N- and C-terminal regions of the other plasmid initiator proteins (Figure 2). Furthermore, the polar amino acids (Lys36, Glu167, Glu26, Gln171, Asn22 and Thr147), which are responsible for interactions between the N- and C-terminal domain, are also conserved in the other plasmid initiator proteins (Figure 2). These conserved polar amino acids could also play a role in maintaining the dual structure.

As mentioned above, the initiator proteins of mini-F-like plasmids are expected to share the dual structure including the HTH motif with RepE. It is also clear from the amino acid sequence alignment that these initiator proteins have a normal HTH motif in the N-terminal region which has the same length as that of RepE (Figure 2). On the other hand, these initiator proteins may have their own unique HTH motif in the C-terminal region with a characteristic size of turn. Therefore, the recognition helix in the C-terminal regions of other plasmid initiator proteins might be used for specific DNA binding. In fact, an HTH motif can be predicted from the amino acid sequence in the corresponding C-terminal region of the RepA initiator of plasmid pPS10, where it plays a crucial role in RepA binding to both operator and iteron sequences (Giraldo *et al.*, 1998). Figure 6C shows a comparison between the cognate DNA binding sites (iteron) of RepE and RepA of pPS10 plasmid. Arg200, Asp203, Arg206 and Arg207 in the C-terminal domain of RepE are responsible for making specific contacts on DNA. The corresponding amino acids of RepA are Asp178, Asp181, Lys184 and Arg185, which are almost similar to those of RepE except for Asp178. It is noteworthy that the DNA bases in the iteron of mini-F plasmid are identical to the corresponding DNA bases in the iteron of pPS10 plasmid, except for guanine which interacts with Arg200. A basic residue of RepE, Arg200, makes contact with the O6 atom of the guanine base. On the other hand, an acidic residue of RepA, Asp178, can form a hydrogen bond with the N4 atom of the cytosine base. The difference between Arg200 of RepE and Asp178 of RepA might explain why these proteins recognize different iteron DNA sequences. The specificity of DNA binding by plasmid initiator proteins may derive from the complementary electrostatic charge distribution of these amino acids in their recognition helices. In addition to the HTH motif, RepE has β -turn- β motifs that also contribute to DNA binding. The Arg-Gly sequences in the β -turn- β region of both domains are conserved in other plasmid initiator proteins (Figure 2). Therefore, the HTH motifs and β -turn- β motifs in both domains probably constitute a common structure necessary for the binding of these plasmid initiator proteins to cognate DNAs (20–22 bp) carrying two major grooves. These results strongly suggest that these plasmid initiator proteins have similar tertiary folding with two structurally similar but functionally different domains.

Comparison with other proteins

For viral origin-binding proteins, crystal structures have been hitherto available only for the DNA-binding domains of E2, EBNA1 and SV40 T-antigen (Hegde *et al.*, 1992; Bochkarev *et al.*, 1996; Luo *et al.*, 1996; Edwards *et al.*, 1998). In eukaryotes, DNA replication is also initiated by the binding of initiator proteins to an origin region on the viral DNA (Edwards *et al.*, 1998). The DNA-binding domains of these viruses have a common structural feature containing a four-stranded antiparallel β -sheet with two α -helices on one side, although they display no amino acid sequence similarities. Interestingly, both EBNA1 and E2 bind to their origins in a remarkably similar dimeric form with the eight-stranded antiparallel β barrel, which comprises four strands from each monomer. Therefore, these proteins seem to share a common ancestor (Edwards

et al., 1998). However, the present RepE54 structure of mini-F plasmid is completely different from the DNA-binding domains of viral origin-binding proteins.

Various DNA binding proteins act as dimers and bind to DNA at two binding sites like virus initiator proteins (EBNA1 and E2). In contrast, RepE binds DNA as a monomer through an internal dual structure comprising two similar binding sites. It has been shown that the TATA-box binding protein (TBP) also has an internal dual structure, which was easily predicted from its primary structure (Nikolov, 1992). The two structural domains of TBP are topologically identical, with a C_{α} r.m.s. deviation of 1.1 Å (Burley, 1996). In contrast, the structural domains of the RepE monomer are not completely identical, and have a C_{α} r.m.s. deviation of 5.1 Å. Furthermore, the N-terminal domain has additional secondary structures (β 2a, β 2b and α 5) that might be responsible for dimerization. This remarkable difference generates the differential binding scheme at the origin and the operator which defines the dual function of RepE.

Unwinding mechanism of DNA duplex for initiation

The most important role of the RepE protein is to induce local melting of the duplex DNA followed by unwinding of the adjacent AT-rich region (13mer region). It is supposed that, in addition to the RepE monomers, the binding of the HU protein is necessary to initiate the local melting of the DNA (Kawasaki *et al.*, 1996). Another essential host factor, DnaA, delivers helicase efficiently to the single stranded region to generate a prepriming complex, but may not be necessary for opening of the DNA duplex. The ability to melt the DNA duplex might be related to the presence of four repeated iteron sequences, each of which can bind a RepE monomer (Figure 7). The spaces between the first and second iterons and between the third and fourth iterons are 2 and 3 bp, respectively. In contrast, there is a 14 bp region between the second and third iterons. This region could provide a binding site for an HU dimer. The origin DNA could be sharply bent at this 14 bp stretch by the binding of HU, and the DNA duplex might be wound around the RepE–HU–origin complex. This model is supported by the observation that the molar ratio of the HU monomer to the RepE monomer required for local melting at the origin of mini-F is 2:4 (Kawasaki *et al.*, 1996). We suppose that the RepE–HU–origin complex can provoke unwinding of the neighboring AT-rich duplex DNA to initiate replication. This may be the reason why the HU protein is indispensable for local melting of the duplex DNA by RepE (Kawasaki *et al.*, 1996). This superhelical DNA model might also be applicable to other iteron-containing replicons in prokaryotes (Bramhill and Kornberg, 1988).

Materials and methods

Crystallization and X-ray data collection

The RepE54 protein was purified as a His-tagged protein by affinity column chromatography after overexpression in *E.coli*. The synthetic DNA oligomers used in complex formation were obtained commercially. Crystals of the RepE54–DNA complex were obtained as previously reported; space group C2 with unit cell parameters of $a = 108.4$ Å, $b = 81.9$ Å, $c = 73.9$ Å and $\beta = 121.5^{\circ}$ (Komori *et al.*, 1999). Intensity data were collected with synchrotron radiation at BL6A and BL18B of

the Photon Factory, KEK (proposal number: 97G095) (Sakabe, 1991) and at BL41XU of SPring-8 (proposal number: 1998A0177-NL) (Kamiya *et al.*, 1995). Several data sets for derivatives were also collected by a RAXIS imaging plate detector with a rotating anode X-ray source. All the intensity data were processed using the program DENZO (Otwinowski and Minor 1997).

MIRAS phasing, model building and refinement

Mercury derivative crystals were prepared by soaking the native crystals in the mother liquid containing 0.1 mM CH_3HgCl for 17 h. IdU derivative crystals were prepared by cocrystallization with iteron DNA substituted with 5-iodouracil for thymine at specific positions (12T, 13T and 33T; Figure 3A). Major heavy-atom sites were determined from difference Patterson maps, and relative positions of the sites between derivatives were determined by the difference Fourier technique using the CCP4 suite (Collaborative Computational Project No. 4, 1994). Multiple isomorphous replacement with anomalous scattering (MIRAS) phase calculations and phase refinements were performed using MLPHARE. The MIRAS phases calculated at 3.0 Å were improved by iterative solvent flattening and histogram mapping with the program DM. The map showed fine electron densities of the protein and DNA molecules with good connectivities. The atomic model was constructed using the graphics program O (Jones and Kjeldgaard, 1994). Crystallographic refinement was carried out by energy minimization and simulated annealing with molecular dynamics using the program X-PLOR (Brünger, 1992a). In the final stage, the native data set was anisotropically scaled to the calculated structure factors of the refinement model, and individual atomic B factors were introduced into the refinement. The atomic model was finally refined to an R -factor of 21.3% (free R -factor = 27.4%; Brünger, 1992b) for 13 738 reflections between 6.0 and 2.6 Å resolution. Analysis of the stereochemistry by PROCHECK showed that all the protein residues were within the allowed regions of the Ramachandran plot (Laskowski *et al.*, 1993).

Acknowledgements

We would like to thank N.Sakabe, N.Watanabe, M.Suzuki and N.Igarashi of the Photon Factory and N.Kamiya, M.Kawamoto and Y.Kawano of SPring-8 for data collection with synchrotron radiation; N.Sasai for initial crystallization; and I.Tanaka, Hokkaido University, for providing the most recent coordinates of HU and critical reading of the manuscript. K.M. is a member of the Sakabe Project of TARA (Tsukuba Advanced Research Alliance), University of Tsukuba. This project was partly supported by the 'Research for the Future' Program from the Japan Society for the Promotion of Science to K.M. (JSPS-RFTF 97L00501), Grant-in-Aid for JSPS Fellows to H.K. (No. 4741) and by Grants-in-Aid for Scientific Research from the Ministry of Education, Science, Sports and Culture, Japan, to C.W.

References

- Armstrong, K.A., Acosta, R., Ledner, E., Machida, Y., Pancotto, M., McCormick, M., Ohtsubo, H. and Ohtsubo, E. (1984) A 37×10^3 molecular weight plasmid-encoded protein is required for replication and copy number control in the plasmid pSC101 and its temperature-sensitive derivative pFS1. *J. Mol. Biol.*, **175**, 331–348.
- Bochkarev, A., Barwell, J.A., Pfuetzner, R.A., Bochkareva, E., Frappier, L. and Edwards, A.M. (1996) Crystal structure of the DNA-binding domain of the Epstein–Barr virus origin-binding protein, EBNA1, bound to DNA. *Cell*, **84**, 791–800.
- Bramhill, D. and Kornberg, A. (1988) A model for initiation at origins of DNA replication. *Cell*, **54**, 915–918.
- Brennan, R.G. (1993) The winged-helix DNA-binding motif: another helix–turn–helix takeoff. *Cell*, **74**, 773–776.
- Brennan, R.G. and Matthews, B.W. (1989) The helix–turn–helix DNA binding motif. *J. Biol. Chem.*, **264**, 1903–1906.
- Brünger, A.T. (1992a) X-PLOR version 3.1: a system for X-ray crystallography and NMR. Yale University Press, New Haven, CT.
- Brünger, A.T. (1992b) The free R value: a novel statistical quantity for assessing the accuracy of crystal structures. *Nature*, **355**, 472–474.
- Bukau, B. and Horwich, A.L. (1998) The Hsp70 and Hsp60 chaperone machines. *Cell*, **92**, 351–366.
- Burley, S.K. (1996) The TATA-box binding protein. *Curr. Opin. Struct. Biol.*, **6**, 69–75.

- Collaborative Computational Project No. 4. (1994) The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D*, **50**, 760–763.
- da Silva-Tatley, F.M. and Steyn, L.M. (1993) Characterization of a replicon of the moderately promiscuous plasmid, pGSH5000, with features of both the mini-replicon of pCU1 and the *ori-2* of F. *Mol. Microbiol.*, **7**, 805–823.
- del Solar, G., Giraldo, R., Ruiz-Echevarria, M.J., Espinosa, M. and Diaz-Orejas, R. (1998) Replication and control of circular bacterial plasmids. *Microbiol. Mol. Biol. Rev.*, **62**, 434–464.
- Edwards, A.M., Bochkarev, A. and Frappier, L. (1998) Origin DNA-binding proteins. *Curr. Opin. Struct. Biol.*, **8**, 49–53.
- Garcia de Viedma, D., Giraldo, R., Rivas, G., Fernandez-Tresguerres, E. and Diaz-Orejas, R. (1996) A leucine zipper motif determines different functions in a DNA replication protein. *EMBO J.*, **15**, 925–934.
- Germino, J. and Bastia, D. (1984) Rapid purification of a cloned gene product by genetic fusion and site-specific proteolysis. *Proc. Natl Acad. Sci. USA*, **81**, 4692–4696.
- Gilbride, K.A. and Brunton, J.L. (1990) Identification and characterization of a new replication region in the *Neisseria gonorrhoeae* β -lactamase plasmid pFA3. *J. Bacteriol.*, **172**, 2439–2446.
- Giraldo, R., Nieto, C., Fernandez-Tresguerres, M.E. and Diaz, R. (1989) Bacterial zipper. *Nature*, **342**, 866.
- Giraldo, R., Andreu, J.M. and Diaz-Orejas, R. (1998) Protein domains and conformational changes in the activation of RepA, a DNA replication initiator. *EMBO J.*, **17**, 4511–4526.
- Hansen, E.B. and Yarmolinsky, M.B. (1986) Host participation in plasmid maintenance: dependence upon *dnaA* of replicons derived from P1 and F. *Proc. Natl Acad. Sci. USA*, **83**, 4423–4427.
- Hegde, R.S., Grossman, S.R., Laimins, L.A. and Sigler, P.B. (1992) Crystal structure at 1.7 Å of the bovine papillomavirus-1 E2 DNA-binding domain bound to its DNA target. *Nature*, **359**, 505–512.
- Ishiai, M., Wada, C., Kawasaki, Y. and Yura, T. (1994) Replication initiator protein RepE of mini-F plasmid: functional differentiation between monomers (initiator) and dimers (autogenous repressor). *Proc. Natl Acad. Sci. USA*, **91**, 3839–3843.
- Jones, T.A. and Kjeldgaard, M. (1994) *O—The Manual*. Uppsala University, Uppsala, Sweden.
- Kamiya, N. *et al.* (1995) Fundamental design of the high energy undulator pilot beamline for macromolecular crystallography at the SPring-8. *Rev. Sci. Instrum.*, **66**, 1703–1705.
- Kawasaki, Y., Wada, C. and Yura, T. (1990) Roles of *Escherichia coli* heat shock proteins DnaK, DnaJ and GrpE in mini-F plasmid replication. *Mol. Gen. Genet.*, **220**, 277–282.
- Kawasaki, Y., Wada, C. and Yura, T. (1991) Binding of RepE initiator protein to mini-F DNA origin (*ori2*): enhancing effects of *repE* mutations and DnaJ heat shock protein. *J. Biol. Chem.*, **267**, 11520–11524.
- Kawasaki, Y., Matsunaga, F., Kano, Y., Yura, T. and Wada, C. (1996) The localized melting of mini-F origin by the combined action of the mini-F initiator protein (RepE) and HU and DnaA of *Escherichia coli*. *Mol. Gen. Genet.*, **253**, 42–49.
- Kline, B.C. (1985) A review of mini-F plasmid maintenance. *Plasmid*, **14**, 1–16.
- Kline, B.C., Kogoma, T., Tam, J.E. and Shields, M.S. (1986) Requirement of the *Escherichia coli dnaA* gene product for plasmid F maintenance. *J. Bacteriol.*, **168**, 440–443.
- Komori, H., Sasai, N., Matsunaga, F., Wada, C. and Miki, K. (1999) Crystallization of replication initiator protein (RepE54) of mini-F plasmid complexed with iteron DNA. *J. Biochem.*, **125**, 24–26.
- Kraulis, P.J. (1991) MOLSCRIPT: a program package to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallogr.*, **24**, 946–950.
- Krishnan, B.R., Fobert, P.R., Seitzer, U. and Iyer, V.N. (1990) Mutations within the replicon of the IncN plasmid pCU1 that affect its *Escherichia coli polA*-independence but not its autonomous replication ability. *Gene*, **91**, 1–7.
- Laskowski, R.J., MacArthur, M.W., Moss, D.S. and Thornton, J.M. (1993) PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.*, **26**, 283–290.
- Luo, X., Sanford, D.G., Bullock, P.A. and Bachovchin, W.W. (1996) Solution structure of the origin DNA-binding domain of SV40 T-antigen. *Nature Struct. Biol.*, **3**, 1034–1039.
- Masson, L. and Ray, D.S. (1986) Mechanism of autonomous control of the *Escherichia coli* F plasmid: different complexes of the initiator/repressor protein are bound to its operator and to an F plasmid replication origin. *Nucleic Acids Res.*, **14**, 5693–5711.
- Matsunaga, F., Kawasaki, Y., Ishiai, M., Nishikawa, K., Yura, T. and Wada, C. (1995) DNA-binding domain of the RepE initiator protein of mini-F plasmid: involvement of the C-terminal region. *J. Bacteriol.*, **177**, 1994–2001.
- Matsunaga, F., Ishiai, M., Kobayashi, G., Uga, H., Yura, T. and Wada, C. (1997) The central region of RepE initiator protein of mini-F plasmid plays a crucial role in dimerization required for negative replication control. *J. Mol. Biol.*, **274**, 27–38.
- Merrit, E.A. and Murphy, M.E. (1994) Raster3D Version 2.0: a program for photorealistic molecular graphics. *Acta Crystallogr. D*, **50**, 869–873.
- Murakami, Y., Ohmori, H., Yura, T. and Nagata, T. (1987) Requirement of the *Escherichia coli dnaA* gene function for *ori2*-dependent mini-F plasmid replication. *J. Bacteriol.*, **169**, 1724–1730.
- Murotsu, T., Matsubara, K., Sugisaki, H. and Takanami, M. (1981) Nine unique repeating sequences in a region essential for replication and incompatibility of the mini-F plasmid. *Gene*, **15**, 257–271.
- Nieto, C., Giraldo, R., Fernandez-Tresguerres, E. and Diaz, R. (1992) Genetic and functional analysis of the basic replicon of pPS10, a plasmid specific for *Pseudomonas* isolated from *Pseudomonas syringae* pathovar *savastanoi*. *J. Mol. Biol.*, **223**, 415–426.
- Nikolov, D.B., Hu, S.H., Lin, J., Gasch, A., Hoffmann, A., Horikoshi, M., Chua, N.H., Roeder, R.G. and Burley, S.K. (1992) Crystal structure of TFIID TATA-box binding protein. *Nature*, **360**, 40–46.
- Ogura, T., Niki, H., Kano, Y., Imamoto, F. and Hiraga, S. (1990) Maintenance of plasmids in HU and IHF mutants of *Escherichia coli*. *Mol. Gen. Genet.*, **220**, 197–203.
- Otwinowski, Z. and Minor, W. (1997) Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.*, **276**, 307–326.
- Sakabe, N. (1991) X-ray diffraction data collection system for modern plate using synchrotron radiation. *Nucl. Instrum. Methods*, **A303**, 448–463.
- Schultz, S.C., Shields, G.C. and Steitz, T.A. (1991) Crystal structure of a CAP–DNA complex: the DNA is bent by 90°. *Science*, **253**, 1001–1007.
- Tanaka, I., Appelt, K., Dijk, J., White, S.W. and Wilson, K.S. (1984) 3 Å resolution structure of a protein with histone-like properties in prokaryotes. *Nature*, **310**, 376–381.
- Wada, M., Kohno, K., Imamoto, F. and Kano, Y. (1988) Participation of *hup* gene product in *ori2*-dependent replication of fertility plasmid F. *Gene*, **70**, 393–397.

Received May 31, 1999; revised and accepted July 14, 1999