Structural Basis for Replication Origin Unwinding by an Initiator Primase of Plasmid ColE2-P9

DUPLEX DNA UNWINDING BY A SINGLE PROTEIN*

Hirosi Itou‡*, Masaru Yagura§, Yasuo Shirakihara‡, and Tateo Itoh§

From the ‡Structural Biology Center, National Institute of Genetics, Yata 1111, Mishima, Shizuoka 411-8540, Japan, the §Department of Cell Genetics, National Institute of Genetics, Yata 1111, Mishima, Shizuoka 411-8540, Japan, and the ¶Department of Biology, Faculty of Science, Shinshu University, Asahi 3-1-1, Matsumoto 390-8621, Japan

Background: Duplex DNA is generally unwound by protein oligomers prior to replication.

Results: The structure of the DNA-binding domain of the ColE2-P9 replication initiation protein bound to Ori DNA was determined.

Conclusion: The ColE2 Rep-DBD unwinds duplex DNA by the concerted actions of its three contiguous structural modules.

Significance: A novel mechanism for duplex DNA unwinding by a single protein was proposed.

Duplex DNA is generally unwound by protein oligomers prior to replication. The Rep protein of plasmid ColE2-P9 (34 kDa) is an essential initiator for plasmid DNA replication. This protein binds the replication origin (Ori) in a sequence-specific manner as a monomer and unwinds DNA. Here we present the crystal structure of the DNA-binding domain of Rep (E2Rep-DBD) in complex with Ori DNA. The structure unveils the basis for Ori-specific recognition by the E2Rep-DBD and also reveals that it unwinds DNA by the concerted actions of its three contiguous structural modules. The structure also shows that the functionally unknown PriCT domain, which forms a compact module, plays a central role in DNA unwinding. The conservation of the PriCT domain in the C termini of some archaeo-eukaryotic primases indicates that it probably plays a similar role in these proteins. Thus, this is the first report providing the structural basis for the functional importance of the conserved PriCT domain and also reveals a novel mechanism for DNA unwinding by a single protein.

DNA replication is the universal process for the transmission of genetic information. To initiate DNA replication, initiator proteins specifically bind to replication origins and unwind limited local regions of replication origins by themselves or with the aid of cooperative proteins. Subsequently, the replication proteins, such as helicase, primase, and DNA polymerase, are loaded onto the unwound region. A number of eukaryotic and prokaryotic initiators, including *Escherichia coli* DnaA, the best characterized replication initiator (1), belong to the superfamily of AAA+ (ATPase associated with diverse cellular activities) proteins, and ATP binding modulates their DNA binding affinity and the formation of higher order structures consisting of multiple proteins (2). The creation of a complex of the ATP-bound DnaA protein oligomer and duplex oriC DNA is critical for local unwinding of an AT-rich DNA unwinding element (DUE)° in oriC and for binding of DnaA protomers to single-stranded DUE (3). Comparative structural studies of the initiator proteins in complex with their cognate origins have suggested certain common mechanistic properties of the engagement with DNA and unwinding of its double-helix structure among the AAA+ initiator proteins (reviewed by Duderstadt and Berger (4)).

The replication initiation of most plasmids requires a specific initiator protein, Rep, encoded by each plasmid. Multiple Rep protomers specifically bind to the cognate origin and then recruit host replication proteins. Unwinding of the origin often relies on host DnaA, and DNA replication initiates more or less similarly to host chromosomal DNA replication (reviewed by del Solar et al. (5)). Several crystal structures of Rep proteins from plasmids harbored in Gram-negative bacteria (F, R6K, and pPS10) and Gram-positive bacteria (pSK41 and pTZ2162) were reported, and these studies revealed the structural bases for their specific binding to DNA and the role of multimer formation in their functions (6–10). The RepE protein of plasmid pAMβ1 from Gram-positive bacteria is exceptional; it specifically binds to the cognate origin and melts DNA duplex as a monomer, although the mechanism is unknown. It has been proposed that additional RepE molecules bind to the melted region to stabilize the unwound DNA (11).

The replication initiator Rep of the *E. coli* plasmid ColE2-P9 (297 amino acids, 34 kDa) is another exception. This Rep specifically binds to the replication origin of the plasmid (Ori; 31 bp) (12–15), and the Ori-bound Rep locally unwinds duplex DNA in the Ori and uniquely exhibits origin-specific primase activity, synthesizing the primer RNA, ppApGpA (16, 17). The primer is used by the host DNA polymerase I to initiate DNA synthesis (16, 18, 19). The ColE2-P9 Ori is the smallest origin...
Structural Basis for Origin Unwinding by ColE2 Rep

among those identified and analyzed to date (15, 20). This Ori can be divided into three functional regions; subregions I and II are important for Rep binding, whereas subregions II and III are important for initiation of DNA replication, and subregion III contains the template sequence for primer RNA synthesis (15) (Fig. 1A). Rep has been thought to consist of two functional domains: an N-terminal primase domain and a C-terminal DNA-binding domain (E2Rep-DBD, residues 175–297) (Fig. 1B). The C-terminal region of E2Rep-DBD determines its binding specificity to Ori (17, 21, 22). The N-terminal region of E2Rep-DBD contains a functionally unknown region called the PriCT (primase C-terminal) domain that is commonly found near the C termini of some primases belonging to the superfamily of archaeo-eukaryotic primases (23). The PriCT domain of E2Rep-DBD has been shown to be involved in origin unwinding (17, 21). Furthermore, Rep binds to and unwinds Ori as a monomer (24). All of these data suggest a novel mechanism of origin unwinding by the E2Rep-DBD. For the past 20 years, however, a growing number of plasmids containing putative Rep proteins and origins closely related to those of plasmid ColE2-P9 have been described (15). These ColE2-related plasmids were found in both Gram-negative and Gram-positive host bacteria and now form a large plasmid family (NCBI nucleotide database). They probably share a common mechanism of plasmid DNA replication initiation that is mediated by initiator proteins closely related to ColE2 Rep (ColE2 Rep-related proteins).

The unique initiation mechanism of the plasmid ColE2-P9 DNA replication has thus far been studied using biochemical and genetic approaches. The molecular mechanisms of the specific binding to and unwinding of Ori by Rep, however, have remained unknown due to the lack of structural information. In this paper, we present the crystal structure of the ColE2 Rep protein in complex with Ori DNA. This structure indicates that the conserved yet functionally unknown PriCT domain plays important roles in Ori unwinding. The results therefore provide the first structural evidence demonstrating the functional importance of the conserved PriCT domain and reveal how a single protein can mediate DNA unwinding.

EXPERIMENTAL PROCEDURES

Sample Preparation—The E2Rep-DBD protein (Arg175–Lys297) was overexpressed in E. coli BL21 star (DE3) cells (Invitrogen) using the pET-28M vector, which is the modified version of the pET-28b(+) vector (Merck Millipore) to attach an octapeptide (MGHHHHHH) at the N terminus of the target protein. The cells were grown at 37 °C until reaching an A600 of 0.5, and overexpression of E2Rep-DBD was induced by adding isopropyl β-D-1-thiogalactopyranoside to a final concentration of 0.5 mM for 10 h at 16 °C before harvesting. The cells were suspended in a buffer containing 20 mM Tris-HCl, pH 7.5, 1.0 M NaCl, 20 mM imidazole, 10% glycerol, 5 mM MgSO4, 1 mM phenylmethylsulfonyl fluoride, and DNase I and disrupted by sonication. The cell lysate was centrifuged to clarify, and the supernatant was applied onto a His-trap crude FF column (GE Healthcare) equilibrated with a wash buffer containing 20 mM Tris-HCl, pH 7.5, 1.0 M NaCl, 20 mM imidazole, and 10% glycerol. The column was washed with 6 column volumes of the wash buffer, and the bound proteins were eluted with 2 column volumes of the elution buffer containing 20 mM Tris-HCl, pH 7.5, 1.0 M NaCl, 0.5 M imidazole, and 10% glycerol. The eluted protein was purified using a Superdex75pg 16/60 size exclusion chromatography column (GE Healthcare) equilibrated with the buffer containing 20 mM Tris-HCl, pH 7.5, 1.0 M NaCl, and 10% glycerol. The purified protein was mixed with 23-bp Ori DNA at a molar ratio of protein/DNA = 1:1.1 for crystallization. The oligonucleotide DNA fragments were purchased from the manufacturer (Hokkaido System Science, Hokkaido, Japan). The protein/DNA mixture was dialyzed against the buffer containing 20 mM Tris-HCl, pH 7.5, 0.1 M NaCl, 10 mM MgCl2, and 1 mM dithiothreitol. Complex formation was confirmed by the electrophoretic mobility shift assay, and the complex was concentrated for crystallization using a centrifugal ultrafiltration device (Sartorius Stedim Biotech GmbH, Göttingen, Germany).

Crystalization—Initial crystallization screening of the E2Rep-DBD-Ori DNA complex was performed using the sitting drop vapor diffusion method in a 96-well crystallization plate (Greiner Bio-One GmbH, Kremsmünster, Austria) at 20 °C with the commercially available screening kits Crystal Screen, Crystal Screen II, and Natrix (Hampton Research, Aliso Viejo, CA) and Wizard I and Wizard II (Emerald BioStructures, Bedford, MA). The crystals used for structure determination were obtained in a solution consisting of 0.1 M sodium citrate, pH 5.4, 0.6 M ammonium sulfate, and 0.5 M LiSO4 in a 24-well VDX plate (Hampton Research) at 20 °C. The crystals took several weeks to grow.

Data Collection and Structure Determination—Native x-ray intensity data were collected from a cryocooled crystal at Beamline BL17A in Photon Factory (KEK, Tsukuba, Japan). The heavy atom derivative was obtained by the soaking method (5 mM K2Pt(CN)4 in the solution). Single wavelength anomalous diffraction data were collected using the platinum-soaked crystal on Beamline AR-NE3A at the Photon Factory. The data were processed using HKL2000 (25). A substructure analysis was performed by the single wavelength anomalous diffraction method using SOLVE (26), and the initial phases were calculated using PHASER (27) in the PHENIX suite (28). The platinum atoms were found in the intermolecular spaces among the protein molecules in the crystal, and the partial atomic model corresponding to the N-terminal domain of the protein was obtained using RESOLVE (29, 30). The map showed electron density corresponding to the DNA backbone and periodic base pairs; thus, the remaining parts of the model were built manually with COOT (31) through iterative model extension and refinement cycles. The overall model was refined using the native data to 2.7 Å resolution using LAFFIRE (32) with REFMAC5 (33) in the CCP4 suite (34). The summaries of data collection and structure analysis are shown in Table 1.

RESULTS

Overall Structure of E2Rep-DBD in Complex with Ori—The crystal structure of the complex of E2Rep-DBD (residues 175–294) and 23-bp DNA comprising subregions I and II of Ori (Fig. 1A) was determined to 2.7 Å resolution. The final atomic model includes E2Rep-DBD residues Arg175–Lys291, the non-template strand (5 Thy–22 Thy), and the template strand (13Gua–32Cyt) (Fig. 2). Helices H1–H3 form a helix bundle, which
module). Helices H6–H7 in this module form an HTH motif with H7 corresponding to the replication helix (Fig. 2A). The three E2Rep-DBD modules are connected with loops and form an elongated fold, indicating that the protein conformation is likely to be flexible, or at least not fixed, without its binding partner, duplex Ori DNA. The surface of E2Rep-DBD is significantly positively charged (Fig. 2B), and nearly one-third of the surface area of the protein (2,606 Å² of 9,583 Å² calculated by using PISA (35)) contributes to its interaction with DNA. Although E2Rep-DBD is small (15 kDa), this interaction surface area is comparable with that observed in larger protein-DNA complexes, such as the archaean DNA replication initiator Cdc6-Orc1 complex (Protein Data Bank entry 2QBY) (36), which has a molecular mass of 85 kDa and a protein-DNA interaction surface area with its 33-bp duplex DNA of 2,699 Å².

The HTH module binds to the major groove near the upstream end of Ori DNA (Figs. 2A and 3). The H4 module binds to the minor groove of the duplex DNA, and the double-helical structure of the DNA is broken near the N terminus of this module. The PriCT module binds to the single-stranded template strand of Ori DNA, just upstream of the template sequence for the primer RNA, whereas the non-template strand extends toward the solvent without any interaction with the protein. The unwinding of a limited region of Ori DNA in the crystal is consistent with previous biochemical data (17).

The crystallographic asymmetric unit contains two complex molecules (complex I: Protein Data Bank chains A, C, and D; complex II: Protein Data Bank chains B, E, and F). Comparison of their tertiary structures showed that they are in slightly different conformations. The structures of the N-terminal segments (amino acid residues 175–239, which correspond to the PriCT module) of the two complex molecules were identical, and so were those of the C-terminal segments (amino acid residues 249–291, which correspond to the H4 and HTH modules). The root mean square deviations for comparisons are 0.27 and 0.54 Å for the main chain atoms of the N- and C-terminal segments of the protein, respectively (the values were calculated using PROLSQ in the CCP4 suite (34)). Obvious differences, however, existed in the relative positions of the two segments in the complexes (Fig. 2C). The overall structure and the protein–DNA interactions, especially in the C-terminal segment, were very similar between these complexes (see below for further discussion). Thus, in this paper, only the structure of complex I has been shown in the figures, except for Figs. 2 (C and D) and 4C.

Flexible yet Locally Stable Structure of E2Rep-DBD Is Capable of Specific Interaction with a Long Region of Double-stranded Ori—E2Rep-DBD has an elongated fold, and thus the HTH and H4 modules can interact with sites of duplex Ori DNA that are considerably distant from each other. The HTH module of E2Rep-DBD exclusively binds to subregion I of Ori, which is important for stably anchoring Rep onto the DNA (15). Five interactions with the bases (A286, G286, G284, and T28) and three interactions with the phosphate backbone (Ser281, Thr284, and Lys289) from helix H7, the DNA-binding helix, were observed (Figs. 3 and 4A). Helix H6, the first helix of the HTH motif, is shorter than those found in canonical HTHs, probably due to...
the presence of Pro\(^{274}\), which is located immediately N-terminal to the helix. As a result, the E2Rep-DBD HTH is more compact than the typical HTH motif. Arg\(^{270}\) and Ser\(^{271}\) in helix H5, adjacent to the HTH motif, also contribute to base recognition and DNA binding stabilization, respectively (Figs. 3 and 4A).

The C-terminal region of Rep containing this HTH module is necessary and sufficient for specific binding to Ori (17). Consistent with these data, the crystal structure clearly shows that the Rep HTH module is the primary region that mediates binding specificity and affinity with its Ori site. It is noted that the tight binding of the HTH module to Ori DNA through these extensive contacts using the amino acids from helices H7 and H5 results in widening of the DNA major groove (17 Å compared with 11 Å in B-DNA).

The H4 module and the loop connecting this module to the C-terminal HTH module (H4–H5 loop) bind along the minor groove of duplex DNA to the upstream and central parts of subregion II (Fig. 3), which is important for binding with Rep and replication (15, 17). Seven interactions with the phosphate backbone (Ser\(^{246}\), Arg\(^{252}\), Arg\(^{254}\), Arg\(^{255}\), Lys\(^{259}\), Lys\(^{261}\), and Arg\(^{262}\)) and one with the base (Gln\(^{249}\)/Gua) were observed. The H4–H5 loop and the C-terminal half of helix H4 have no base-specific interactions with the upstream half of subregion II (Figs. 3 and 4B). This part of Ori contains AT-rich sequences, which are known to be intrinsically bent (37). In fact, the duplex DNA bends around the upstream end of the AT tract (14Ade–17Ade) by 31°. There is also a slight twist along its helical axis by 6° toward duplex unwinding with no obvious widening of the minor groove (analyzed by using CURVES+ (38)) (Fig. 2D).

SELEX experiments using Rep gave rise to a consensus sequence corresponding to subregions I and II up to the 20Cyt/18′Gua pair of Ori (39), and the substitution of each base pair in
the AT tract to a GC pair decreased the binding affinity by Rep (15). These results showed that Rep prefers DNA fragments containing the specific AT-rich sequence located in the upstream region of subregion II. Therefore, the local conformation of duplex DNA, which depends on the nucleotide sequence, contributes to the binding specificity and affinity of the H4 module to Ori, although no base-specific interactions are observed in this region in the crystal structure. An amino acid substitution at the glutamine involved in the sole base-specific interaction observed in the H4 module (Q249R) abolished the Rep DNA replication activity (40), and all nucleotide substitutions at the 20Cyt/18Gua pair of Ori recognized by Gln249 severely decreased its DNA binding affinity (15). All of these results indicated the involvement of the H4 module in specific binding with Ori.

The H4 module Has a Critical Role in Site-specific Duplex DNA Unwinding—The duplex structure of DNA bound by Rep is highly distorted near the N-terminal part of the Rep H4 module. DNA strands are twisted by 20° (complex II) or 25° (complex I) along its helical axis at the 20Cyt/18Gua pair, and the downstream strands are separated (Figs. 2 and 3). Phe245 and Val248 in the H4 module, along with Phe240 in the PriCT module, stack with the 16Ade base in the single-stranded template strand. The side chain of Gln249 in the helix H4 interacts with the 18Gua base in the template strand (Fig. 4B), as mentioned above, and the 20Cyt/18Gua pair is a pivotal point of the conformational change that occurs in the duplex DNA (Fig. 2C).

A deletion derivative of Rep containing only the H4 and HTH modules induced weak but significant distortion of duplex DNA at the 24Thy/14′Ade and 25Cyt/13′Gua pairs in subregion II (17, 40), indicating the importance of the specific binding of the H4 module to Ori, including the specific interaction of Gln249 with 18′Gua to trigger unwinding. The H4 module plays an important role in site-specific unwinding of Ori, in cooperation with the PriCT module (see below).

The PriCT Module Specifically Binds to the Single-stranded Template DNA for Primer RNA Synthesis—The Rep-bound duplex DNA is unwound in the AT-rich region (21Thy/17′Ade to 24Thy/14′Ade) around the downstream end of subregion II of Ori (Figs. 2A and 3). The aromatic side chains of Rep residues Phe180 and Tyr235 stack with the 13′Gua and 14′Ade bases in the single-stranded template strand, respectively, and additionally, Asn176, Ser231, and the hydroxyl group of Tyr188 form hydrogen bonds with the 13′Gua and 14′Ade bases, respectively (Figs. 3 and 4C). Phe240 stacks with the 16′Ade base along with Phe245 and Val248 in the H4 module, as mentioned above. Arg184, Arg189, and Arg192 in the PriCT module contact the phosphate backbone of the single-stranded template strand on the surface of the module, and Lys239 interacts with the 15′Thy base in complex I (Fig. 4C).

Mutational analyses showed that single substitutions of the amino acids interacting specifically with the bases (N176S, N176D, F180S, F180L, Y188N, Y188H, S231N, and F240S) resulted in loss of DNA replication activity (40). Among the ColE2-related Rep proteins (Fig. 5A), Asn176, Phe180, Tyr188, Ser231, and Phe245 are perfectly conserved, Tyr235 and Val248 are conserved or changed conservatively, and Phe240 is conserved in many of the ColE2-related Rep proteins. All of these facts support the importance of the tight binding between the PriCT module and Ori DNA in the initiation of plasmid DNA replication.

The PriCT Module Shifts Its Position at the Loop Connecting This Module to the Contiguous Modules—Comparison of the two complexes in the crystallographic asymmetric unit showed that E2Rep-DDB in complex with Ori DNA exhibits structural flexibility around the loop connecting the PriCT and H4 modules (H3-H4 loop). In complex I, the PriCT module shifts its position (Fig. 2C), and the single-stranded template strand bound by the PriCT module is more distorted from the duplex form than in complex II. An amino acid substitution of Ser241 in the H3-H4 loop to proline resulted in a loss of DNA replication activity (40). Such a substitution causes a structural constraint in the loop connecting the PriCT and H4 modules, supporting the importance of structural flexibility between the modules. The conformational change causes a slight difference in the protein-DNA interactions. The most prominent difference was the AT tract to a GC pair decreased the binding affinity by Rep (15).
observed in the interaction between Lys$_{239}$ and 15'Thy (Fig. 4C). The torsion angle of the single-stranded template strand in complex I is different from that in complex II due to the conformational change, and as a consequence, the orientation of the 15'Thy base differs between the complexes. The distance between the Ne atom of Lys$_{239}$ and the O4 atom of 15'Thy is 2.7 Å in complex I, whereas that in complex II is 5 Å. Thus, the side chain of Lys$_{239}$ forms a hydrogen bond with 15'Thy in complex II.
I to stabilize the conformation, whereas it cannot do so in complex II (Fig. 4C). The side chain of Tyr235, which stacks with 14/11032 Ade, shifts slightly, and Val248 stacking with 16/11032 Ade alters its position due to the conformational change between the modules. These changes lead to increased contact area with the bases in complex I, suggesting that this complex is more stable than complex II. The combined results suggest that unwinding of the Ori may occur in association with the shift in the position of the PriCT module relative to the other modules, and the two conformations in the crystal structure possibly reflect the structural changes that occur during duplex DNA unwinding.

**DISCUSSION**

Our structure of the E2Rep-DDBD bound to Ori DNA showed that the Rep protein unwinds duplex DNA by the concerted actions of three contiguous domains, the HTH-, H4-, and PriCT modules. The locally folded structure of each module and flexibility of the regions connecting them seem to be important for main-
Structural Basis for Origin Unwinding by ColE2 Rep

FIGURE 6. Model for unwinding of the complete Ori by Rep. The protein and DNA strands are shown with the same color code as in Fig. 1A. The surface charge representation is as shown in Fig. 2B. The green ribbon shows the protein main chain, and sky blue and yellow space-filling models show the basic and aromatic residues conserved around the basic channel, respectively. Yellow and magenta, bases of the nucleotides in the non-template and the template strands, respectively. Black and gray dotted lines, predicted trajectories of DNA strands and the linker connecting helix H1 and the N-terminal primase domain, respectively. Red dotted oval, region of Rep important for the interaction with the single-stranded non-template strand (40).

Possible Interaction of the PriCT module with the Non-template Strand to Stabilize the Unwound Conformation—In the crystal structure, the single-stranded non-template strand does not interact with the protein. The surface of the PriCT module near the 3'-end of the unwound non-template strand is highly basic, and a closer look at the crystal structure revealed that there is also a basic channel on its surface, which is likely to bind to the non-template DNA strand (Fig. 2B). The basic amino acids Arg175, Lys182, Arg185, Arg189, and Arg208 and the aromatic amino acids Trp186 and Tyr212 are located along the channel. These amino acids are conserved among most of the ColE2 Rep-related proteins (Fig. 5A), and substitutions of the completely conserved Arg175 and Trp186 residues of Rep (R175Q and W186R, respectively) resulted in loss of DNA replication activity (40), indicating the importance of these amino acids. It has also been shown that the intact Rep protein has stronger affinity for the single-stranded non-template strand than for the single-stranded template strand. The region from Arg165 to Leu179 of Rep, the N-terminal part of the PriCT module and the C-terminal half of the presumed linker region connecting the PriCT module to the N-terminal primase domain (Figs. 1B and 5A), is important for the interaction (40). The region is also important for unwinding of Ori around the template sequence for primer RNA synthesis and the adjacent downstream AT-rich region (26Ade/12Thy to 31Ade/7Thy). These combined data suggest that the interaction of the non-template strand with the N terminus of the PriCT module and the linker connecting this module to the N-terminal primase domain promotes and stabilizes further unwinding of the Ori around the template sequence for primer RNA synthesis (Fig. 6). The structural disorder of the non-template strand in the crystal may be explained if the downstream region of Ori, which

taining stable and specific binding to Ori and unwinding of its duplex conformation. Although the tertiary structure of E2RepDBD is rather unusual as an initiator, its conformation is quite suitable for its functions of specific binding to Ori and maintaining strong affinity to distort its duplex conformation.

Importance of the Local Distortion, Shift in Position of the PriCT module, and Supercoiling for Unwinding Duplex Ori DNA—Unwinding of duplex DNA at replication origins by initiator proteins generally depends on DNA supercoiling (41, 42). Rep unwinds the AT-rich region (21Thy/17Ade to 30Thy/8Ade) of Ori in a negative supercoiling-dependent manner (40). In addition, Rep induces distortion of duplex DNA at the 24Thy/14Ade and 25Cyt/13Gua base pairs, albeit very weakly, even when the DNA used is linear. A deletion derivative of Rep containing only the H4 and HTH modules also induces weak distortion of duplex DNA at these pairs (17), as mentioned above. These results indicated that the 24Thy/14Ade and 25Cyt/13Gua pairs are more or less distorted by the interaction with the H4 module, independently of the DNA conformation. Such a local distortion of duplex DNA may trigger unwinding of Ori in the adjacent AT-rich region, driven by the intrinsic torsional strain of DNA negative supercoiling. Then interaction of the PriCT module with bases 13’Gua and 14’Ade promotes unwinding. Structural flexibility between the H4 and PriCT modules and the shift in position of the PriCT module relative to the other modules (the conformational change from complex II to complex I) might be important for unwinding, as mentioned above. In the crystal structure, the 24Thy/14Ade and 25Cyt/13Gua pairs are located at the ends of the oligonucleotide DNA used, so unwinding is possible only by binding of the PriCT module to these pairs and by free rotation of the duplex DNA without any contribution from DNA negative supercoiling.
is absent from the oligonucleotide DNA used in this study, is required to maintain a stable interaction with the PriCT module. Because the 13’Gua and 14’Ado bases are specifically and tightly bound by the PriCT module and the template region for primer RNA synthesis (10’Thy-12’Thy) are located immediately downstream of these bases (Fig. 6), the PriCT module is likely to function as the platform stabilizing the template strand during subsequent sequence specific primer synthesis by the N-terminal primase domain.

**Functional Implication for the Other ColE2 Rep-related Proteins**—The crystal structure of E2Rep-DBD in complex with Ori provides insights into the specificity determination mechanism between Rep and Ori in ColE2-related plasmids. Previous studies showed that a critical specificity determinant for stable binding of Rep to Ori between the ColE2-P9 and ColE3-CA38 plasmids is located in the region B and site β (Fig. 5, A and B), and Thr284 in helix H7 of ColE2 Rep and its corresponding residue Thr293 of ColE3 Rep are the key residues in binding (21, 22). The crystal structure showed that ColE2 Rep Thr284 contacts the phosphate backbone of Gua in the non-template strand, whereas Arg287 in the same helix makes specific contacts with the Gua base (Figs. 3 and 4A). The crystal structure suggests that the substitution of the amino acid in position 284 from threonine to tryptophan would hamper the interaction between Arg287 and Gua through steric hindrance by the bulky tryptophan side chain. Nucleotide substitutions at Gua significantly decreased the binding affinity to Rep (15), and the R287Q mutation abolishes the DNA replication ability of Rep (40), indicating the importance of the Arg287-Gua interaction for stable binding of Rep and Ori in plasmid ColE2-P9. In the case of plasmid ColE3-CA38, the Arg287 is replaced with glutamine; thus, its base recognition differs from that of plasmid ColE2-P9. ColE3 Rep has lysine and arginine at positions 289 and 291 of ColE2 Rep (Fig. 5A), and these basic residues are conserved among the Rep proteins that contain tryptophan at position 284 of ColE2 Rep (ColE3-CA38, ColE5-099, and pEI2 Rep proteins in Fig. 5A). These observations implied the involvement of these basic residues in specificity determination, and in fact, studies using the mutated Rep proteins showed some contribution of the conserved basic residues for the specificity determination between ColE2-P9 and ColE3-CA38 plasmids (21). The present crystal structure showed that the side chain of the lysine is likely to contact only the DNA phosphate backbone, and the arginine side chain is too distant to interact with the paired bases. Because several important amino acids of ColE2-P9 Rep that are involved in its interaction with Ori are changed in these Rep proteins, the relative position of helix H7, the DNA-binding helix, to the DNA major groove might be slightly shifted for specific binding.

The Rep proteins of some of the ColE2-related plasmids have insertions or deletions of amino acids on either or both sides of the conserved Gln249 in the H4 module (regions A and C in Fig. 5A), and their Ori regions correspondingly have insertions or deletions of nucleotides at either or both sides of the conserved Cyt/Gua pair (sites α and γ in Fig. 5B). Regions A of Rep and the sites α of Ori are also the determinants of plasmid specificity between ColE2-P9 and ColE3-CA38 plasmids and involved in site-specific unwinding of the origin regions (21, 22). By modulating the structures of the H4 modules of the Rep proteins, the spacings between the H4 modules (especially the conserved glutamines) and the HTH module (region A and site α) and PriCT modules (region C and site γ) could be adjusted to place the three modules at their proper positions on the cognate Ori regions, which is important for specific binding, duplex DNA unwinding, and subsequent primer RNA synthesis. This may well be the case for the other ColE2-related plasmids.

Many of the amino acids in the H4 and PriCT modules of Rep involved in base-specific interactions and stabilization of the complex are highly conserved in many of the ColE2 Rep-related proteins (Fig. 5A). The sequences of subregions II and III of the origins of the ColE2-related plasmids are nearly identical; the regions bound by the PriCT module are particularly conserved (22Thy–25Cyt) (Fig. 5B). These findings strongly suggest that the mechanism of Ori unwinding around the primer template sequence by the Rep proteins is shared by the ColE2-related plasmids.

**PriCT Domain, a Novel Protein Module for Duplex DNA Unwinding**—The primases belonging to some clades of archaeo-eukaryotic primases commonly contain either one of the two conserved, distantly related α-helical modules called PriCT-1 and PriCT-2 domains (23). The PriCT module of ColE2 Rep belongs to the PriCT-1 domain, and bioinformatic studies showed that PriCT-1 domains are located C-terminal to the primase domains in all of the archaeo-eukaryotic primases (23), just as in ColE2 Rep. This suggests that the PriCT-1 domains of the archaeo-eukaryotic primases also form compact modules and serve as platforms for primer RNA synthesis. The PriCT-2 domains could have a similar structure and function. This is the first report providing a structural basis to demonstrate the importance of the conserved but, until now, functionally uncharacterized PriCT domain (23).

The RepE of plasmid pAMβ1 binds to the cognate origin and melts DNA duplex as a monomer (11); thus, it would be reasonable to assume that ColE2 Rep and pAMβ1 RepE share a certain common functional property. A closer look at the amino acid sequence revealed that the pAMβ1 RepE also contains a region showing some homology with the PriCT domain in ColE2 Rep (Fig. 5A). Asn176, Phe180, Phe245, and Gln249 in ColE2 Rep, which directly interact with the bases, are perfectly conserved, whereas the other important residues Tyr235 and Phe240 are conservatively changed and Tyr188 and Ser231 are varied (Fig. 5A). This possibly reflects the difference in the binding sequences of DNA between pAMβ1 RepE and ColE2 Rep. The involvement of the PriCT homology region of pAMβ1 RepE in origin melting is strongly suggested, although its molecular mechanism remains unclear.

**Comparison with Other Rep Proteins**—The crystal structures of Rep proteins belonging to the other Rep family, RepA of the plasmids pPS10 and pR6K, and RepE of F from Gram-negative bacteria have already been solved (6–9). These Rep proteins commonly use a winged HTH motif for their binding with DNA. Although the detailed mechanism is still unclear, it has been proposed that the conformational change of DNA duplex induced by binding of multiple Rep molecules onto the repeated iterons, their binding sites in Ori, promotes recruitment of host replication proteins (6, 8, 43). Recently, the struc-
Structural Basis for Origin Unwinding by ColE2 Rep

tural study of RepA of pSK41 and pZ2162 from Gram-positive bacteria has been reported (10). Although the amino acid sequences of Gram-positive RepA are dissimilar from those of Gram-negative ones, their DNA-binding domains contain winged HTH as a DNA-binding motif. The crystal structure showed that the DNA-binding domain of RepA binds to DNA as a dimer, and the wing region distorts DNA duplex at the minor groove. Because multiple RepA dimers were predicted to bind to Ori from the same side of the duplex DNA molecule, a model in which the local distortion of the DNA duplex is amplified by the cooperative interaction of the multiple RepA dimers was suggested. Structural similarity between RepA and a host replication protein was observed; thus, the involvement of RepA in subsequent primosome assembly is also suggested (10).

As described above, ColE2 Rep is unique because it binds to Ori and unwinds its duplex structure as a monomer without the aid of any host replication protein. The tertiary structure of Rep is also rather unusual as an initiator protein but is suitable for exhibiting its functions. The PriCT module, the conserved compact helical module, plays a central role for DNA duplex unwinding. Because the ColE2 Rep-related proteins form a large protein family, the results and discussion described here will accelerate our understanding of the initiation mechanism of DNA replication, a common fundamental biological process of every organism in the transmission of genetic information, through providing mechanistic insight into origin DNA unwinding for one of the diversified groups of plasmid initiator proteins.

Acknowledgments—We thank the staff of beamlines BL17A and BL-NE3A for help with diffraction data collection. We thank Dr. H. Araki (National Institute of Genetics) and Dr. I. Tanaka (Hokkaido University) for discussions and critical comments. We thank the anonymous referees for the careful reading of the manuscript and the valuable suggestions to improve it.

REFERENCES

1. Katayama, T., Ozaki, S., Keyamura, K., and Fujimitsu, K. (2010) Regulation of the replication cycle: conserved and diverse regulatory systems for DnaA and oriC. Nat. Rev. Microbiol. 8, 163–170
2. Neuwald, A. F., Aravind, L., Spouge, J. L., and Koonin, E. V. (1999) AAA+: A class of chaperone-like ATPases associated with the assembly, operation, and disassembly of protein complexes. Genome Res. 9, 27–43
3. Ozaki, S., and Katayama, T. (2012) Highly organized DnaA-oriC complexes recruit the single-stranded DNA for replication initiation. Nucleic Acids Res. 40, 1648–1665
4. Duderstadt, K. E., and Berger, J. M. (2013) A structural framework for replication origin opening by AAA+ initiation factors. Curr. Opin. Struct. Biol. 23, 144–153
5. del Solar, G., Giraldo, R., Ruiz-Echevarria, M., Espinosa, M., and Diaz-Orejas, R. (1998) Replication and control of circular bacterial plasmids. Microbiol. Mol. Biol. Rev. 62, 434–464
6. Komori, H., Matsunaga, F., Higuchi, Y., Ishiai, M., Wada, C., and Miki, K. (1999) Crystal structure of a prokaryotic replication initiator protein bound to DNA at 2.6 Å resolution. EMBO J. 18, 4597–4607
7. Nakamura, A., Wada, C., and Miki, K. (2007) Structural basis for regulation of bifunctional roles in replication initiator protein. Proc. Natl. Acad. Sci. U.S.A. 104, 18484–18489
8. Swan, M. K., Bastia, D., and Davies, C. (2006) Crystal structure of pi initiator protein-iteron complex of plasmid R6K: implications for initiation of plasmid DNA replication. Proc. Natl. Acad. Sci. U.S.A. 103, 18481–18486
9. Giraldo, R., Fernández-Tornero, C., Evans, P. R., Díaz-Orejas, R., and Romero, A. (2003) A conformational switch between transcriptional repression and replication initiation in the RepA dimerization domain. Nat. Struct. Biol. 10, 565–571
10. Schumacher, M. A., Tonthat, N. K., Kwong, S. M., Chinnam, N. B., Liu, M. A., Skurray, R. A., and Firth, N. (2014) Mechanism of staphylococcal multiresistance plasmid replication origin assembly by the RepA protein. Proc. Natl. Acad. Sci. U.S.A. 111, 9121–9126
11. Le Chatelier, E., Jannière, L., Ehrlich, S. D., and Cancelier, D. (2001) The RepE initiator is a double-stranded and single-stranded DNA-binding protein that forms an atypical open complex at the onset of replication of plasmid pAMB1 from Gram-positive bacteria. J. Biol. Chem. 276, 10234–10246
12. Itoh, T., and Horii, T. (1989) Replication of ColE2 and ColE3 plasmids in vitro replication dependent on plasmid-coded proteins. Mol. Gen. Genet. 219, 249–255
13. Kido, M., Yasueda, H., and Itoh, T. (1991) Identification of a plasmid-coded protein required for initiation of ColE2 DNA replication. Nucleic Acids Res. 19, 2875–2880
14. Yasueda, H., Horii, T., and Itoh, T. (1989) Structural and functional organization of ColE2 and ColE2 replisomes. Mol. Gen. Genet. 215, 209–216
15. Yagura, M., Nishio, S. Y., Kurozumi, H., Wang, C. F., and Itoh, T. (2006) Anatomy of the replication origin of plasmid ColE2-P9. J. Bacteriol. 188, 999–1010
16. Takechi, S., Matsu, H., and Itoh, T. (1995) Primer RNA synthesis by plasmid-specific Rep protein for initiation of ColE2 DNA replication. EMBO J. 14, 5141–5147
17. Han, M., Yasuda, M., and Itoh, T. (2007) Specific interaction between the initiator protein (Rep) and origin of plasmid ColE2-P9. J. Bacteriol. 189, 1061–1071
18. Takechi, S., and Itoh, T. (1995) Initiation of unidirectional ColE2 DNA replication by a unique priming mechanism. Nucleic Acids Res. 23, 4196–4201
19. Kingsbury, D. T., and Helinski, D. R. (1970) DNA polymerase as a requirement for the maintenance of the bacterial plasmid colicinogenic factor E1. Biochem. Biophys. Res. Commun. 41, 1580–1544
20. Horii, T., and Itoh, T. (1988) Replication of ColE2 and ColE3 plasmids: the regions sufficient for autonomous replication. Mol. Gen. Genet. 212, 225–231
21. Aoki, K., Shinohara, M., and Itoh, T. (2007) Distinct functions of the two specificity determinants in replication initiation of plasmids ColE2-P9 and ColE3-CA38. J. Bacteriol. 189, 2392–2400
22. Shinohara, M., and Itoh, T. (1996) Specificity determinants in interaction of the initiator (Rep) proteins with the origins in the plasmids ColE2-P9 and ColE3-CA38 identified by chimera analysis. J. Mol. Biol. 257, 290–300
23. Iyer, L. M., Koonin, E. V., Leipe, D. D., and Aravind, L. (2005) Origin and evolution of the archaeo-eukaryotic primase superfamily and related palm-domain proteins: structural insights and new members. Nucleic Acids Res. 33, 3875–3896
24. Han, M., Aoki, K., Yasuda, M., and Itoh, T. (2007) The ColE2-P9 Rep protein binds to the DNA as a monomer. Biochem. Biophys. Res. Commun. 353, 306–310
25. Otwowski, Z., and Minor, W. (1997) Processing of x-ray diffraction data collected in oscillation mode. Methods Enzymol. 276, 307–326
26. Terwilliger, T. C., and Berendzen, J. (1999) Automated MAD and MIR structure solution. Acta Crystallogr. D Biol. Crystallogr. 55, 849–861
27. McCoy, A. J., Gross-Kuleste, B. W., Adams, P. D., Winn, M. D., Storoni, L. C., and Read, R. J. (2007) Phaser crystallographic software. J. Appl. Crystallogr. 40, 658–674
28. Adams, P. D., Fonine, P. V., Bunkóczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L. W., Kapral, G. J., Gross-Kuleste, R. W., McCoy, A. J., Moriarty, N. W., Oeffner, R., Read, R. J., Richardson, D. C., Richardson, J. S., Terwilliger, T. C., and Zwart, P. H. (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta Crystallogr. D Biol. Crystallogr. 66, 213–221
29. Terwilliger, T. C. (2000) Maximum likelihood density modification. Acta Crystallogr. D Biol. Crystallogr. 56, 965–972
30. Terwilliger, T. C. (2003) Automated main-chain model-building by tem-
plate-matching and iterative fragment extension. Acta Crystallogr. D Biol.
Crystallogr. 59, 38–44
31. Emsley, P., and Cowtan, K. (2004) Coot: model-building tools for molec-
ular graphics. Acta Crystallogr. D Biol. Crystallogr. 60, 2126–2132
32. Yao, M., Zhou, Y., and Tanaka, I. (2006) LAFFIRE: software for automating
the refinement process of protein-structure analysis. Acta Crystallogr. D
Biol. Crystallogr. 62, 189–196
33. Murshudov, G. N., Vagin, A. A., and Dodson, E. J. (1997) Refinement of
macromolecular structures by the maximum-likelihood method. Acta Crystallogr.
D Biol. Crystallogr. 53, 240–255
34. Collaborative Computational Project, Number 4 (1994) The CCP4 suite:
programs for protein crystallography. Acta Crystallogr. D Biol. Crystallogr.
50, 760–763
35. Krissinel, E., and Henrick, K. (2007) Inference of macromolecular assem-
blies from crystalline state. J. Mol. Biol. 372, 774–797
36. Dueber, E. L., Corn, J. E., Bell, S. D., and Berger, J. M. (2007) Replication
origin recognition and deformation by a heterodimeric archaeal Orc1
complex. Science 317, 1210–1213
37. Carrera, P., and Azorín, F. (1994) Structural characterization of intrinsi-
cally curved AT-rich DNA sequences. Nucleic Acids Res. 22, 3671–3680
38. Lavery, R., Moakher, M., Maddocks, J. H., Petekeviciute, D., and Zakrze-
ska, K. (2009) Conformational analysis of nucleic acids revisited: Curves+
Nucleic Acids Res. 37, 5917–5929
39. Yagura, M., and Itoh, T. (2006) The Rep protein binding elements of the
plasmid ColE2-P9 replication origin. Biochem. Biophys. Res. Commun.
345, 872–877
40. Han, M. (2007) Molecular mechanisms of early steps of ColE2 plasmid
replication initiation. Ph.D. thesis, Shinshu University, Matsumoto, Japan
41. Boulakas, T. (1996) Common structural features of replication origins in
all life forms. J. Cell. Biochem. 60, 297–316
42. Zannis-Hadjopoulos, M., and Price, G. B. (1999) Eukaryotic DNA replica-
tion. J. Cell. Biochem. (Suppl. 32/33), 1–14
43. Giraldo, R., and Fernández-Tresguerres, M. E. (2004) Twenty years of the
pPS10 replicon: insights on the molecular mechanism for the activation of
DNA replication in iteron-containing bacterial plasmids. Plasmid 52,
69–83
44. Baker, N. A., Sept, D., Joseph, S., Holst, M. J., and McCammon, J. A. (2001)
Electrostatics of nanosystems: application to microtubules and the ribo-
some. Proc. Natl. Acad. Sci. U.S.A. 98, 10037–10041
45. Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan,
P. A., McWilliam, H., Valentin, F., Wallace, I. M., Wilm, A., Lopez, R.,
Thompson, J. D., Gibson, T. J., and Higgins, D. G. (2007) Clustal W and
Clustal X version 2.0. Bioinformatics 23, 2947–2948
46. Laskowski, R. A., Mac Arthur, M. W., Moss, D. S., and Thornton, J. M.
(1993) PROCHECK: a program to check stereochemical quality of protein
structure. J. Appl. Crystallogr. 26, 283–291