Structural basis of the 3'-end recognition of a leading strand in stalled replication forks by PriA

Kaori Sasaki¹, Toyoyuki Ose¹, Naoaki Okamoto², Katsumi Maenaka¹, Taku Tanaka³, Hisao Masai³, Mihoko Saito⁴, Tsuyoshi Shirai⁴ and Daisuke Kohda^{1,*}

¹Division of Structural Biology, Medical Institute of Bioregulation, Kyushu University, Fukuoka, Japan, ²Olympus Corp., Tokyo, Japan, ³Genome Dynamics Project, Tokyo Metropolitan Institute of Medical Science, Tokyo, Japan and ⁴Department of Bioscience, Nagahama Institute of Bioscience and Technology, and JST-BIRD, Siga, Japan

In eubacteria, PriA helicase detects the stalled DNA replication forks. This critical role of PriA is ascribed to its ability to bind to the 3' end of a nascent leading DNA strand in the stalled replication forks. The crystal structures in complexes with oligonucleotides and the combination of fluorescence correlation spectroscopy and mutagenesis reveal that the N-terminal domain of PriA possesses a binding pocket for the 3'-terminal nucleotide residue of DNA. The interaction with the deoxyribose 3'-OH is essential for the 3'-terminal recognition. In contrast, the direct interaction with 3'-end nucleobase is unexpected, considering the same affinity for oligonucleotides carrying the four bases at the 3' end. Thus, the N-terminal domain of PriA recognizes the 3'-end base in a base-non-selective manner, in addition to the deoxyribose and 5'-side phosphodiester group, of the 3'-terminal nucleotide to acquire both sufficient affinity and non-selectivity to find all of the stalled replication forks generated during DNA duplication. This unique feature is prerequisite for the proper positioning of the helicase domain of PriA on the unreplicated double-stranded DNA. *The EMBO Journal* (2007) 26, 2584–2593. doi:10.1038/sj.emboj.7601697; Published online 26 April 2007

Subject Categories: genome stability & dynamics; structural biology

Keywords: PriA; primosome; protein–nucleic acid interaction; stalled DNA replication fork; X-ray crystallography

Introduction

When cells are stressed by ultraviolet irradiation or other damaging treatments, DNA replication forks often encounter template DNA lesions, which can stall the progression of the replication forks (Kowalczykowski, 2000). Processing of the stalled replication forks by recombination proteins, including RecA, RecG, and other Rec and Ruv proteins, is necessary in *Escherichia coli* cells to form intermediate DNA structures such as a chicken-foot fork structure and a D-loop (displace-

ment loop) structure (Kowalczykowski, 2000). These branched DNA structures are the target of two proteins, PriA (McGlynn *et al.*, 1997; Liu and Marians, 1999; Marians, 2000; Xu and Marians, 2003) and PriC (North and Nakai, 2005; Heller and Marians, 2005a). The two proteins ultimately load the DnaB replicative helicase and initiate the reassembly of a replisome to continue the duplication of genomic DNA. PriA was originally identified as a protein essential for the replication of the small, single-stranded DNA (ssDNA) of the phage ϕ X174 (Shlomai and Kornberg, 1980). PriA functions as a scaffold that recruits other proteins, including PriB, PriC, DnaT, DnaC, DnaB, and DnaG. The protein cluster is collectively referred to as the ϕ X174-type primosome. Previous studies clarified that the mechanism and the order of the primosome assembly (Ng and Marians, 1996; Jones and Nakai, 1997; Liu and Marians, 1999). In addition to the phage replication, PriA is necessary for the induction of RecA-dependent stable DNA replication (Masai *et al.*, 1994), normal cell proliferation, and resistance to various genotoxic agents, including UV and mitomycin C (Lee and Kornberg, 1991; Kogoma *et al.*, 1996).

PriA contains a DEXH-type DNA 3' → 5' helicase domain in the C-terminal region (Ouzounis and Blencowe, 1991). Despite the conservation of the helicase motifs among eubacteria, the helicase activity of the C-terminal domain is unexpectedly dispensable for most of the PriA functions (Zavitz and Marians, 1992; Liu *et al.*, 1999), with the exception of its requirement for full-level DNA synthesis in recombination-dependent modes of DNA replication (Tanaka *et al.*, 2003). Recently, the PriA helicase activity was reported to prevent RecA from provoking unnecessary recombination during replication fork restarting (Mahdi *et al.*, 2006), and to unwind the nascent lagging strand DNA to help PriC load the DnaB helicase onto stalled replication forks where there is a gap in the nascent leading strand (Heller and Marians, 2005b), although another helicase, Rep, can substitute for the function of the PriA helicase in both cases.

Replication fork reassembly requires the recognition of a specific DNA structure by PriA. PriA reportedly binds to a D-loop-like structure by recognizing a bend at the three-strand junction, and to duplex DNA with a protruding 3' single strand (McGlynn *et al.*, 1997; Nurse *et al.*, 1999; Chen *et al.*, 2004), but the entity that PriA recognizes was unclear. We discovered that the N-terminal 181-residue domain of *E. coli* PriA specifically bound to a D-loop structure and an A-fork structure in gel shift assays (Tanaka *et al.*, 2002; Mizukoshi *et al.*, 2003), and to a ssDNA in a surface plasmon resonance (SPR) analysis (Mizukoshi *et al.*, 2003). In both experiments, the phosphorylation of the 3'-hydroxyl group of the 3'-terminal nucleotide residue decreased the binding substantially, suggesting that the N-terminal domain of PriA recognizes an unmodified 3' end of ssDNA. It is reasonable that PriA monitors the occurrence of an unusual 3' end of DNA as a landmark to detect the stalling of chromosomal DNA replication. Recently, we showed that the presence of a

*Corresponding author. Division of Structural Biology, Medical Institute of Bioregulation, Kyushu University, Maidashi 3-1-1, Higashi-ku, Fukuoka 812-8582, Japan. Tel.: +81 92 642 6968; Fax: +81 92 642 6764; E-mail: kohda@bioreg.kyushu-u.ac.jp

Received: 17 November 2006; accepted: 29 March 2007; published online: 26 April 2007

free 3' end at the branch point of an arrested fork structure and the 3'-end-binding ability of PriA were both required for the stabilization of the arrested fork structure (Tanaka and Masai, 2006). The binding of the N-terminal domain to the 3' end of the leading DNA strand is required to place the helicase domain in an 'unwinding-deficient' orientation. Otherwise, the helicase destabilizes the fork structure and the repair of the arrested replication fork is aborted.

In the present study, we analyzed the interaction between the N-terminal 105-residue domain of PriA and oligonucleotides carrying different bases at the 3' end by fluorescence correlation spectroscopy (FCS) and crystallography. The results clearly indicated that PriA interacts with the base, deoxyribose, and 5'-side phosphodiester group of the 3'-terminal nucleotide residue. The deoxyribose moiety, in particular, the 3'-OH group is recognized by a network of polar interactions. By contrast, the direct recognition of the base moiety was really unexpected, considering the nearly equal affinity of PriA for ssDNA, irrespective of the base at the 3' end. After a systematic mutagenesis study, we concluded that *E. coli* PriA utilizes a unique amino acid-nucleotide interaction mode to acquire sufficient affinity for the 3' end of DNA, in a manner that neither discriminates the 3'-terminal base nor disturbs the Watson-Crick base pairing. The unique 3'-terminal nucleotide recognition is prerequisite for the proper sensing of the stalled replication forks by the PriA protein.

Results

The N-terminal 181-residue domain of PriA consists of two subdomains

We prepared the N-terminal 181-residue fragment of *E. coli* PriA and measured the NMR spectrum (Supplementary Figure 1A). The good dispersion of the cross peaks in the [¹H, ¹⁵N]HSQC spectra indicates its stable tertiary structure, that is, the 181-residue fragment is a structural domain. The fragment 1–181 was cleaved after Arg105 and Arg108 by a limited proteolysis with trypsin (data not shown). We made two new constructs, 1–105 and 109–181, and found that both fragments had a stable fold by NMR (Supplementary Figure 1B and C). Thus, the fragment 1–181 consists of two subdomains, residues 1–105 and residues 109–181. The NMR spectrum of PriA[1–181] roughly equals the sum of the two spectra of PriA[1–105] and of PriA[109–181], suggesting the independency of the two domains.

The N-terminal domain of PriA recognizes the 3'-terminal nucleotide residue

The interaction of the N-terminal 105-residue domain of *E. coli* PriA (PriA[1–105]) with oligodeoxyribonucleotides was analyzed using FCS. FCS is a spectroscopic technique for studying molecular interactions in solution (Rigler, 1995). FCS monitors stochastic spontaneous fluctuations of fluorescently labeled particles, due to their entrance into and exit from a defined, small volume irradiated by a focused laser beam. The correlation function analysis of the fluorescent intensity provides the diffusion time, an average time for the particle to cross the small volume. The diffusion time is dependent on the mass of the particle, and thus the increase in the diffusion time of a fluorescently labeled molecule indicates an interaction with an added protein.

First, we showed the specific recognition of the 3'-terminal nucleotide residue by PriA[1–105] (Figure 1A). As the PriA[1–105] protein was added to 5'TAMRA-d(A₈) (octadeoxyadenylate with a terminal 5' fluorophore and an unmodified 3'-hydroxyl group), a dose-dependent increase in the diffusion time was observed, due to complex formation. The curve fitting revealed a dissociation constant of about 25 μM. By contrast, neither 5'TAMRA-d(A₈)p (terminal 3' phosphate) nor d(A₈)-3'TAMRA (terminal 3' fluorophore) increased the diffusion time at all, indicating that the modification of the 3'-hydroxyl group by a bulky group completely blocks the interaction of PriA[1–105] and the oligonucleotide. The longer version of the N-terminal fragment, PriA[1–181], and the full-length PriA[1–732] bound to the 5'TAMRA-d(A₈) with higher affinities of 7 and 3 μM, respectively, but the helicase domain of PriA[194–732] did not (Figure 1B). We further analyzed the role of the 3'-hydroxyl group in the recognition by changing the 3'-terminal deoxyribose to a dideoxyribose and prepared 5'TAMRA-d(A₇)ddC. This oligonucleotide ends with a dideoxycytidine, and thus lacks hydroxyl groups at the 3' end. The binding was abolished (Figure 1C), indicating that a hydrogen bond involving the 3'-hydroxyl group is essential.

Then, we analyzed the interaction with oligonucleotides of differing lengths, ranging from 2 to 8 nt (Figure 1D). The optimal length was around 4 nt, but the shortest, 2 nt, can still bind to PriA[1–105] with comparable affinity. This result strengthens the notion that PriA[1–105] mainly recognizes the 3'-terminal nucleotide of DNA, and the other parts of the DNA may assist in the binding.

Finally, we analyzed the interaction with oligonucleotides bearing different bases at the 3' end (Figure 1E). We prepared 5'TAMRA-d(A₇C), 5'TAMRA-d(A₇G), and 5'TAMRA-d(A₇T) in addition to 5'TAMRA-d(A₈). PriA[1–105] bound to the four oligonucleotides with almost the same affinity. This intriguing result is not an artifact of domain excision, as the full-length PriA had the same base-non-selectivity (Figure 1F).

Structure determination and overview

The structure of PriA[1–105] in the absence of a ligand was determined using the multiwavelength anomalous diffraction (MAD) method to a resolution of 2.5 Å (Table I). The asymmetric unit in the crystal contains four PriA[1–105] molecules. Overall, the four PriA[1–105] structures in the crystal are very similar, with Cα r.m.s.d. values in the range of 0.8 to 1.3 Å. In the crystal, two PriA[1–105] molecules exchange their N-terminal seven amino-acid residues and form an intertwined dimer (Supplementary Figure 2), and thus there are two independent, intertwined dimers in the asymmetric unit. The PriA[1–105] protein also forms a stable dimer in solution, as demonstrated by gel filtration and analytical ultracentrifugation (data not shown). SPR and NMR analyses in the previous study (Mizukoshi *et al*, 2003), and FCS data (Figure 1) in the present study clearly showed that dimer formation by PriA[1–105] in solution does not interfere with the 3'-terminal nucleotide binding of the oligonucleotides, and that the binding properties of PriA[1–105] are similar to those of the full-length PriA.

The functional form of the N-terminal 105-residue fragment should be monomeric because the full-length PriA is monomeric. We reconstituted the monomeric form of PriA[1–105] by swapping the coordinates of the N-terminal segments between the two intertwined chains and used it in

Figure 1 FCS analyses of PriA binding to TAMRA-labeled oligonucleotides. (A) Plots of FCS diffusion time against protein concentration. The points are connected by solid lines to aid in visualization. The N-terminal domain of PriA (PriA[1-105]) was mixed with 5'TAMRA-d(A₈), 5'TAMRA-d(A₈)p, and d(A₈)-3'TAMRA. (B) Same as in (A) for the full-length PriA (PriA[1-732]), the longer version of the N-terminal domain (PriA[1-181]) and the helicase domain (PriA[194-732]) mixed with 5'TAMRA-d(A₈). (C) Effect of the deletion of the 3'-OH group of the 3'-terminal deoxyribose moiety on binding. PriA[1-105] was mixed at a final concentration of 100 μM with 5'TAMRA-d(A₇)ddC (dideoxycytidine at the 3' end). 5'TAMRA-d(A₇C) was used as a positive control and 5'TAMRA-d(A₇C)p was a negative control. The corrected diffusion time (marked by *) was calculated by subtracting the value at zero protein concentration. (D) Effect of the oligonucleotide length on binding. PriA[1-105] was mixed at a final concentration of 250 μM with 5'TAMRA-d(A_n), where the number of residues, *n*, varied from 2 to 8. (E) Base preference at the 3' end of oligonucleotides. PriA[1-105] was mixed at a final concentration of 250 μM with 5'TAMRA-d(A₇N), where N is A, C, G, or T. 5'TAMRA-d(A₈)p was used as a negative control. (F) Same as in (E) for the full-length PriA[1-732] at a final concentration of 100 μM.

the following discussion (Figure 2A). PriA[1-105] is composed of two antiparallel β -sheets, a one-turn 3_{10} helix, and two α -helices connected by a short loop. The secondary structure with the multiple alignment of the N-terminal domain of PriA from representative bacterial species is shown in Figure 2B. The architecture of the N-terminal domain of PriA appears to be novel, as no equivalent fold was found in a DALI search of the Protein Data Bank (Holm and Sander, 1998).

3'-Terminal nucleotide-binding site

We carried out complex structure determinations of PriA[1-105] with four dinucleotides, d(AA), d(AC), d(AG), and d(AT) (Table I). The dinucleotides used were designed to have a constant adenine base at the 5' end and a variable base at the 3' end for analyzing the 3'-terminal base recognition by PriA[1-105]. In all of the complexes, the ligand binding did

not change the overall protein conformation, with r.m.s.d. values in the range of 0.19 to 0.80 Å.

Not all of the binding sites of PriA[1-105] were occupied by the ligands (Table I). In the crystal structure with d(AA), the electron density of the ligand existed in all of the four independent copies of the protein, but in the structures with the other dinucleotides, the densities of only two or three sites in the four copies were adequate for building nucleotide models, and the low occupancy and/or the high temperature factor prevented us from building a model at the remaining sites. The small variation in the affinity due to subtle differences in the crystallographic environments might explain the reason why not all binding sites are equally occupied. Even at the visible sites, the electron density corresponding to the constant 5'-terminal adenylate was poorly observed, and only that of the 3'-terminal nucleotide was positioned on the

Figure 2 Structure and multiple sequence alignment of the N-terminal domain of PriA. **(A)** The functional monomeric form of the N-terminal 105-residue domain of *E. coli* PriA was reconstituted from the intertwined dimeric form in crystal (Supplementary Figure 2). The swapped N-terminal seven-residue segment is drawn blue. The hinge region (Leu8–Arg14), for which the structure was not determined, is shown as an orange, broken curve. In a DALI search of the Protein Data Bank (Holm and Sander, 1998), the proteins showing the highest structural similarity had z scores in the range of 2.1 to 2.8 and r.m.s.d. values of 2.7–4.3 Å for 43–70 paired C α atoms. **(B)** The secondary structure is shown above the alignment. β , α , T, and η represent a β -strand, an α -helix, a turn, and a 3_{10} helix, respectively. Residues involved in the binding of the 3'-terminal nucleotide in *E. coli* PriA are labeled with blue triangles. Representative sequences were selected from each clade of the phylogeny of eubacteria (Supplementary Figure 6). The figure was produced with ESPript 2.2 (<http://esprict.ibcp.fr/ESPript/ESPript/index.php>).

surface of the PriA[1–105] protein (Figure 3A). All of these observations were attributed to the fact that PriA[1–105] only recognizes the 3'-terminal nucleotide portion (base, deoxyribose, and 5'-side phosphodiester group), with a weak binding affinity. We also prepared a cocrystal of PriA[1–105] with a trinucleotide, d(CCC). As expected, just the density corresponding to the 3'-terminal cytidylate was visible. Comparisons of the structure with d(CCC) to those with d(AA) and d(AC) demonstrated that no serious biases arose in the binding mode from the different lengths of the ligands or the different base next to the 3'-terminal nucleotide.

The crystal structure of PriA[1–105] in complexes with oligonucleotides facilitated the identification of the residues that contact the 3' end of the DNA. PriA[1–105] possesses a

pocket, consisting of Phe16, Asp17, Tyr18, Gly37, Leu55, and Lys61, to accommodate the 3'-terminal nucleotide of an oligonucleotide (Figure 3B). In accordance, the backbone amide cross peaks of Asp17, Tyr18, and Gly37 were the most affected during the titration of an 8-nt oligonucleotide in the previous NMR study (Mizukoshi *et al.*, 2003). Figure 3D summarizes the binding mode schematically. The oxygen atoms of the 5'-side phosphodiester group are liganded to the Gly37 backbone N and the Lys61 side-chain N. The deoxyribose moiety is anchored to the bottom of the binding pocket, by a stacking interaction on Tyr18, and in contact with the Lys61 backbone N and the Asp17 backbone O. Finally, the base is situated in a cleft formed by Phe16, Asp17, and Leu55 and adopts the *anti* conformation about

Table 1 Refinement statistics

Ligand	Free PriA[1–105]	PriA[1–105] complex with				
		d(AA)	d(AC)	d(AG)	d(AT)	d(CCC)
Resolution range (Å)	20.0–2.52	20.0–3.30	20.0–3.20	20.0–3.35	20.0–3.15	20.0–3.0
No. reflections (work/test)	18 496/1418	8466/633	8987/685	8152/646	9894/763	11 132/859
$R_{\text{work}}/R_{\text{free}}$	24.9/29.5	26.0/30.9	24.7/30.2	28.1/33.3	27.1/31.8	23.8/29.3
<i>Number of copies in AU</i>						
Protein	4	4	4	4	4	4
Ligand	0	4	2	2	3	1
<i>Number of atoms</i>						
Protein	3296	3267	3296	3296	3296	3217
Ligand	—	84	77	38	78	19
Water	56	—	—	—	—	—
<i>B-factors</i>						
Protein	74.82	94.968	87.962	108.36	104.01	89.458
Ligand	—	148.61	135.83	132.14	130.94	149.59
<i>R.m.s. deviations</i>						
Bond lengths (Å)	0.008	0.005	0.010	0.011	0.009	0.009
Bond angles (deg)	1.5	1.2	1.7	1.7	1.9	1.6
<i>Ramachandran plot</i>						
Most favored (%)	94.3	80.7	85.2	88.6	87.5	90.7
Additionally allowed (%)	4.5	19.2	13.6	11.4	12.5	7.0
Generously allowed (%)	1.1	0	1.1	0	0	2.3
Disallowed (%)	0	0	0	0	0	0
PDB entry	2D7E	2D7G	2DWL	2DWN	2DWM	2D7H

Each data set was collected from a single crystal. Data collection statistics were reported previously (Sasaki *et al*, 2006).

the glycosyl bond. The base is sandwiched between Phe16 and Leu55, whereas the minor groove face of the base interacts with Asp17 (Figure 3C). Note that the major and minor groove faces of the bases are defined as the directions accessible from the outside in the duplex DNA structure (see Figure 1 of the literature; Luscombe *et al*, 2001).

Effects of point mutations on binding affinity and specificity

We used FCS analysis to monitor the binding of PriA[1–105] mutants to octadeoxyadenylate with a 5'-end fluorophore. We observed a pronounced reduction in the binding affinities by the alanine-scanning mutations of the amino-acid residues in contact with the 3'-terminal nucleotide: Phe16, Asp17, Gly37, Leu55, and Lys61 (Figure 4). To check the structural integrity of the mutants, we measured 1D ^1H NMR spectra (Supplementary Figure 3). The chemical shifts of the upfield shifted methyl peaks, which are highly sensitive to the tertiary fold, are unchanged upon mutations. The mutation of Phe36, next to Gly37, also impaired the binding. As a control, the mutation of Lys38 had little effect on the binding affinity. Notably, the mutation of Tyr18 was not tested because of inclusion body formation, suggesting the important role of Tyr18 in the PriA structure.

The crystal structures suggested that the backbone N and side-chain O of Asp17 are involved in the interaction with the 3'-terminal base. We generated six mutants of Asp17 that differ in charge and size, and monitored the binding specificity for the 3'-terminal base. We used four 5'TAMRA-labeled octanucleotides, each with one of the four nucleotides at the 3' end, in the FCS analysis (Figure 5A). The substitution of Asp17 to any other amino acid always negatively affected the affinity. At the same time, the substitution of Asp17 makes

PriA[1–105] have distinctive specificity for the 3'-terminal base. A subtle change, such as the extension of the length of the side chain of Asp17 by a methylene group (Asp→Glu) or the amidation of the carboxyl group (Asp→Asn), causes PriA[1–105] to have base specificity.

We also prepared mutations of Leu55, which hydrophobically contacts the 3'-terminal base (Figure 5B). The substitution of Leu55 to Val reduced the binding affinity moderately, and that to Ala showed a more pronounced reduction of affinity. The substitution to Glu abolished the binding (data not shown). Interestingly, no obvious differences in the affinity for the 3'-terminal base were generated by the Leu55 mutations, in sharp contrast to the case of the Asp17 mutations.

Discussion

The crystal structure of PriA[1–105] in complexes with di- and trinucleotides facilitated the identification of residues that contact the 3' end of DNA (Figure 3). The electron density of the 3'-terminal nucleotides is of moderate quality. We consider that this is due to the inherent nature of the loose binding mode of PriA. To allow judging the relevance of the models, the ligand-unbiased $F_o - F_c$ maps, calculated from the protein coordinates before the nucleotide coordinates were added during refinement, are shown in Supplementary Figure 4. In addition, the $2F_o - F_c$ maps and the $F_o - F_c$ omit map, both calculated from the final coordinates, are shown in and Figure 3A and Supplementary Figure 4, respectively.

In parallel with the crystal screening of PriA[1–105], we attempted to crystallize the PriA[1–181], which has three times higher affinity for the oligonucleotides, but we have

Figure 3 3'-Terminal nucleotide-binding site of PriA with bound ligands. (A) Stereoview of the electron density of the $F_o - F_c$ omit map calculated from the final coordinates of the PriA[1–105]·d(AA) complex structure, minus the coordinates of the ligand, contoured at 3σ (green). The final refined structure is superimposed on the map. (B) Superimposed five complex structures with d(AA) (green), d(AC) (cyan), d(AG) (magenta), d(AT) (yellow), and d(CCC) (salmon). Only the 3'-terminal nucleotides were visible in the electron density maps and drawn. Amino-acid residues in contact with the 3'-terminal nucleotides are labeled. (C) Different view of the binding site, showing the interactions between the 3'-terminal base, Asp17, and the loop segment, Ser53–Glu54–Leu55. (D) Schematic representation of the interactions at the 3'-terminal nucleotide-binding site in the LIGPLOT style. The 5'-side phosphodiester group, the deoxyribose moiety, and the nucleobase of the 3'-terminal nucleotides are recognized by PriA. Amino-acid residues involved in hydrophobic contacts with the nucleotide are drawn schematically as brown spoked arcs. (E) Site of the 3'-nucleotide binding in the overall structure.

not obtained any crystals, probably due to the flexibility between the two subdomains.

Mechanism for the 3'-terminal nucleotide recognition

The modification of the 3'-OH group of the deoxyribose by a bulky group, such as phosphate or the TAMRA dye, completely blocked the interaction (Figure 1A). The removal of the 3'-OH group also abolished the binding (Figure 1C). Thus, the size and hydrogen bonding at the 3' position of deoxyribose are crucial for the 3'-end recognition by the PriA protein. The stacking interaction between the deoxyribose ring and the tyrosine residue at position 18 also appears important because Tyr18 is one of the small number of invariably conserved amino-acid residues in the PriA protein family (Figure 2B). PriA interacts with the 5'-side phosphodiester

group using the side-chain amino group of a lysine residue and the backbone amide of a glycine residue, but few interactions exist with the 5' neighboring nucleotide residue in all of the crystal structures determined (Figure 3). The extensive interactions of Asp17 and Lys61 with the 3'-terminal nucleotides make the 3'-terminal nucleotide-binding site of PriA[1–105] narrow and compact. In summary, PriA has a small pocket that accommodates the 3'-terminal nucleotide only. In fact, the extension of oligonucleotides enhanced the affinity, but the effect was limited (Figure 1D).

Unique mechanism for the base-non-selective recognition

To achieve base non-selectivity, no specific interactions with bases is a common mechanism. For example, the flap

structure-specific endonuclease FEN-1 binds to the 3' flap (unpaired 3' DNA end) without base specificity and excises the 5' flap. The 3' flap base has no contacts in the structure (Chapados *et al*, 2004). The exosome RNase PH core is a 3'-end RNA exonuclease without base specificity. The 3'-end base stacks nonspecifically on another base and a tyrosine residue (Lorentzen and Conti, 2005). We have found that PriA[1–105] utilizes a distinct mechanism in which the 3' base is specifically recognized by an aspartate residue (Figure 6). One critical contact is a hydrogen bond formed between the backbone NH of Asp17 and the purine base N3 or pyrimidine base O2 atom. Since the N3 and O2 atoms of the bound bases are located at similar positions, this interaction offers a structural basis for the unbiased preference for large purine and small pyrimidine bases. In the case of the pyrimidines, a second hydrogen bond (water mediated) to the

Asp17 side-chain O may form. Additional base-specific, polar interactions with the Asp17 side-chain O are probable for the adenine N1, the guanine N1 and N2, the cytosine N3, and the thymine N3. The interaction between Asp17 and the 3'-terminal base is fine-tuned, because the substitution of Asp17 to other any amino acids made PriA selective to 3'-terminal bases (Figure 5A). The mutations we generated had a tendency for the binding affinity to decrease in the order cytosine, thymine, adenine, and guanine. By the loss of optimized interactions between the base and the Asp17 side chain by the mutations, the common interaction mode involving the backbone NH of Asp becomes predominant. This interaction is probably more favorable for pyrimidine bases than purine bases. The sandwiching of the 3'-terminal base between two hydrophobic residues, Phe16 and Leu55 (Figure 3C), appears to be crucial for the precise positioning of the 3'-terminal base to the aspartate residue at position 17 to fulfill the bipartite interaction between the Asp17 and the 3'-terminal base (Figure 6).

Interchangeable Thr and Ser in addition to Asp at position 17

Given the essential role of the side chain of Asp at position 17 in the *E. coli* PriA structure, the moderate level of conservation (42%, after correction for the biased sampling) in the multiple sequence alignment of the 181 PriA proteins in the HOGENOM database looks unconvincing at first glance. Interestingly, besides Asp, Thr (31%) and Ser (13%) are frequently found, resulting in a high level of combined conservation (86%) at this position. In contrast, the quite rare occurrence of Glu (<1%) clearly indicates that a negative charge is not necessary. This is not an artifact due to misalignment of the sequences, because position 17 is next to the well-conserved Tyr18 residue (Figure 2B). Note that the

Figure 4 FCS analyses of the binding of alanine-scanning mutants of PriA[1–105] to octadeoxyadenylate. Amino-acid residues that contact the 3'-terminal nucleotide in the crystal structures were substituted to an alanine. Wild type and alanine point mutants of PriA [1–105] were mixed at a final concentration of 500 μM with 5'TAMRA-d(A₈). As a negative control, the wild-type PriA[1–105] was mixed with 5'TAMRA-d(A₈)p.

Figure 5 FCS analyses of the binding of mutant PriA[1–105] to oligonucleotides bearing different 3'-terminal bases. (A) Point mutants with the substitution of Asp17 with various amino acids were mixed at final concentrations of 50–100 μM with 5'TAMRA-d(A₇N), which possesses different bases, A, C, G, or T, at the 3' end. (B) Same as in (A), except for point mutants with the substitution of Leu55.

Figure 6 Feasible recognition model for the base non-selective recognition by PriA. The orange dotted line represents a hydrogen bond between the backbone NH of Asp17 and the purine base N3 or pyrimidine base O2 atom of the 3'-terminal nucleotide. This interaction offers a structural basis for the base-non-selective recognition. The other interactions, depicted as green dotted lines, are also involved in the unbiased preference for different nucleobases.

changes from Asp to Thr and Ser arise from double base substitutions.

We postulate that the unique bipartite interaction of Asp17 with the 3'-terminal base explains the unusual group conservation (Figure 6). The simultaneous polar interactions of the backbone N and the side-chain O of Asp17 with the 3'-terminal base constrain the length of the side chain must be in the appropriate range. Namely, Thr and Ser (three bonds between N and O^γ), and Asp (four bonds between N and O^γ) are suitable for the bipartite interaction, but Glu and Gln (five bonds between N and O^ε) are not. Unfortunately, this explanation is not valid for the exclusion of Asn (four bonds between N and O^δ). We surveyed bipartite amino acid–nucleobase interactions similar to Asp17 of PriA in the Protein Data Bank. We found the preference of Asp, Thr, and Ser and the avoidance of Glu, Gln, and Asn at the minor groove face of bases as compared with simple interaction modes at the minor groove face of bases (Supplementary Figure 5). Thus, the uncommon conservation of Asp, Thr, and Ser at position 17 of PriA proteins is ascribed to the geometric requirement of the bipartite interaction with a nucleobase.

We then carried out the distribution analysis of the four PriA types (Asp, Thr, Ser, and other amino acids at position 17) in the phylogenetic tree of eubacteria (Supplementary Figure 6). The multiple, independent occurrences of different types of PriA during evolution suggests that there should be functional necessity for the conservation of Asp, Thr, and Ser at the position 17. We suggest that Thr and Ser, in addition to Asp, contribute to the base non-selectivity. In fact, in *E. coli* PriA, the substitution to Thr generated the least base preference among the six mutations in the present study, although the affinity was reduced (Figure 5A). The biochemical and structural studies of PriA proteins having Thr or Ser at position 17 are necessary to clarify the roles of these residues in the base-non-selective recognition.

Functional implications of the 3'-end recognition by PriA

We previously showed in gel shift assays that PriA bound to the 3' end of the leading strand at the branch point in model DNA structures mimicking a fork and a D-loop (Tanaka *et al*, 2002; Mizukoshi *et al*, 2003). In these structures, the 3'-terminal nucleotide is probably base paired with the complementary nucleotide on the template strand. As the interaction of Asp17 with a base occurs at the minor groove

Figure 7 Schematic of the PriA protein–D-loop DNA structure. The existence of the free 3' end of DNA is unusual in cells, and thus can be a landmark of the stalled DNA replication fork. The interaction of the N-terminal domain of PriA with the 3'-terminus of the leading DNA strand directs the proper positioning of the helicase domain of PriA on the unreplicated double-stranded DNA (Tanaka and Masai, 2006). The N-terminal domain of PriA possesses a small binding pocket consisting of three subsites for the 5'-side phosphate group, the deoxyribose moiety including the 3'-OH group, and the base of the 3'-terminal nucleotide residue of the nascent leading strand. PriA recognizes the four different bases at the 3' end in a non-discriminating manner, without disturbing the base pairing with the template DNA strand. This unique feature enables PriA to find all of the stalled replication forks generated during DNA duplication, with sufficient affinity.

face of the bases (Figure 3), it is rational that the interaction does not prevent the base pairing.

Fork and D-loop structures are generated at any location along the chromosomal DNA, and so the base composition at the 3' end of the nascent leading DNA strand is not necessarily biased. Thus, it is reasonable that PriA is designed to have equal affinities for the four nucleobases at the 3' end. Indeed, PriA recognizes the 3'-terminal base in a base-non-selective manner, in addition to the deoxyribose moiety and 5'-side phosphodiester group, to bind the 3'-terminal nucleotide residue of DNA. Why does PriA recognize the bases to achieve the equal affinities for different 3'-terminal bases? Leaving the 3'-terminal base unattended seems like an easier solution. Probably, the binding affinity with the deoxyribose and phosphodiester group would be insufficient without the interaction with the base. In this context, we predict that an

as yet unidentified functional counterpart of PriA in eukaryotic cells should utilize the Asp/Thr/Ser-minor groove interaction to detect the 3' end in stalled DNA replication forks.

The unique 3'-terminal nucleotide recognition of the N-terminal domain of PriA directs the proper positioning of the helicase domain of PriA on the unreplicated double-stranded DNA in the intermediate DNA structures (Figure 7) (Tanaka and Masai, 2006), which leads to the assembly of the primosome, and eventually restarts the chromosomal DNA duplication.

Materials and methods

Protein purification and crystallization

The N-terminal domain of *E. coli* PriA, encompassing amino acids Met1–Arg105 (PriA[1–105]) (accession number P17888), and its mutants were expressed as N-terminal hexahistidine-tagged proteins in *E. coli*, and were purified as previously described (Sasaki *et al*, 2006). For the FCS analysis, the histidine tag was retained, whereas, for crystallization, the tag was removed by thrombin treatment. After thrombin digestion, the protein had three extra residues (Gly–Ser–His) at the N terminus. The purified protein was concentrated in 20 mM sodium phosphate buffer, pH 5.5, containing 150 mM KCl and 3 mM DTT. For the determination of the concentration of the PriA[1–105] protein, an extinction coefficient of 1.26 at 280 nm was used. A longer version of the N-terminal domain PriA[1–181], the helicase domain PriA[194–732], and the full-length PriA[1–732] were prepared in the same manner as PriA[1–105].

Oligonucleotide synthesis

Oligonucleotides used for crystallization were di- and trioxynucleotides, d(AA), d(AC), d(AG), d(AT), and d(CCC). For the FCS measurement, oligodeoxynucleotides, d(A)_n, *n* = 2–8, d(A₇C), d(A₇G), and d(A₇T), were synthesized with a TAMRA (5/6-carboxy tetramethylrhodamine) fluorophore modification at the 5' end and with an optional 3'-terminal phosphate group (Genenet, Fukuoka, Japan). An oligonucleotide terminating in dideoxyribose at the 3' end, d(A₇)ddC (i.e. dideoxycytidine), was synthesized by Sigma Genosys (Ishikari, Hokkaido, Japan). Cation exchange of oligonucleotides was performed with AG50W resin (BioRad) from the H⁺ to the Na⁺ form.

Structure determination

The crystallization and preliminary crystallographic analysis were previously reported for free, native PriA[1–105], a selenomethionyl (SeMet) derivative of PriA[1–105], and cocrystals with d(AA) and d(CCC) (Sasaki *et al*, 2006). Cocrystals of PriA[1–105] with d(AC), d(AG), and d(AT) were obtained by following the same procedure. The crystals grew at 20°C after 1–3 days, against a well solution of 0.1 M sodium citrate (pH 3.6–4.2) and 0.1–0.3 M ammonium sulfate. For data collection, the crystals were flash frozen (100 K) in the reservoir solution containing 30–40% (v/v) ethylene glycol or glycerol. Diffraction data sets were collected at SPring8 BL38B1 (Harima, Japan) and at Photon Factory BL6A (Tsukuba, Japan). All crystals belonged to the space group R32 (Sasaki *et al*, 2006). The diffraction data were integrated, merged, and scaled with HKL2000 (Otwinowski and Minor, 1997).

We initially solved the structure of *E. coli* PriA[1–105] in the absence of a ligand using the MAD method. Four of the eight Se sites were determined with SOLVE (Terwilliger and Berendzen, 1999), followed by the refinement of heavy-atom parameters and the initial phase calculations using SHARP (de La Fortelle and Bricogne, 1997). The four NCS operators were generated from the partial model, and the phases from SHARP were improved by NCS averaging and solvent flattening with the phase extension procedure to 2.8 Å (native crystal) using DM (Cowtan, 1994). The initial model was built using O (Jones *et al*, 1991) from the results of the automatic model building programs, RESOLVE (Terwilliger, 2002) and ARP/warp (Perrakis *et al*, 1999). Subsequent refinement was carried out using O and CNS (Brunger *et al*, 1998). The final model of the free form includes four PriA[1–105] molecules and 56 water

molecules, and was refined against data from the native crystal to 2.5 Å resolution to an *R*-factor of 24.9% and an *R*_{free} factor of 29.5%. The extra three residues at the N termini were missing and residues 1 and/or 105 were missing for some chains in the final model. The manual building of the models of the structures of PriA[1–105] complexed with di- and trioxynucleotides was accomplished using O. Refinement of the models was carried out using CNS. Refinement statistics are summarized in Table I. The coordinates have been deposited in the Protein Data Bank, with accession codes 2D7E (PriA[1–105]), 2D7G (PriA[1–105]·d(AA)), 2DWL (PriA[1–105]·d(AC)), 2DWN (PriA[1–105]·d(AG)), 2DWM (PriA[1–105]·d(AT)), and 2D7H (PriA[1–105]·d(CCC)). Figures were prepared using PyMOL (<http://pymol.sourceforge.net/>).

Fluorescence correlation spectroscopy

PriA binding of TAMRA-labeled oligonucleotides was analyzed using the single molecule fluorescence detection system MF20 (Olympus, Japan). The FCS sample (20 µl) contained 40 mM HEPES buffer, pH 7.6, 5 mM EDTA, 40 mM glutamate, 1 mM DTT, 20 µg/ml BSA, 2 nM TAMRA-labeled oligonucleotide, and various concentrations (0.1–500 µM, as indicated in the figure legends) of the PriA protein, in a 384-well glass-bottomed microplate. For the detection of TAMRA-labeled molecules, a He–Ne laser (543 nm, 200 µW) was used. The data acquisition time was 15 s per measurement at ambient temperature. The measurement was repeated three times per sample and the average and standard deviation of the diffusion time were calculated. We used several MF20 devices with different laser-irradiated volumes, so the measured diffusion time varied with respect to each device. We always included positive and negative controls in every experiment for comparison.

Systematic search of the interacting amino-acid residues with the bound nucleotides

The Protein Data Bank (Apr 2006 release, 33 595 entries) was used to retrieve protein crystal structures in complexes with DNA and RNA. The initial set contained 1412 entries. Then, the structures determined to a resolution better than 2.5 Å or the structures determined by NMR, if no high-resolution X-ray structure was available, were selected, and redundant structures with homologous sequences were excluded, as previously reported (Saito *et al*, 2006). A total of 2396 nucleotide-binding sites, in 212 structures, were retained in the final set. We searched amino-acid residues in which polar side chain is located within 3.5 Å of any of the purine N3, C2 (Ade), N2 (Gua), or pyrimidine O2 atom to find residues located at the minor groove face of the bases. From the resultant data set, we selected amino-acid residues of the PriA-type bipartite interaction: the backbone N is located within 3.5 Å of purine N3 or pyrimidine O2 atom, and simultaneously the polar side chain is within 6.5 Å, considering water-mediated interactions, of any of the polar atoms of nucleobases.

Multiple sequence alignment

The multiple sequence alignment of the PriA protein was retrieved from the HOGENOM database (release 03, October, 2005; <http://pbil.univ-lyon1.fr/databases/hogenom.html>). The gene family alignment HBG016564 contains 181 sequences. The frequency of the appearance of each amino acid at a given position was calculated using the program 'Site-Specific Res. Frq.' at the Ash server (<http://timpani.genome.ad.jp/~ash/>) for the correction of a biased sampling of genomes.

Supplementary data

Supplementary data are available at *The EMBO Journal* Online (<http://www.embojournal.org>).

Acknowledgements

We thank the staff of beamlines BL38B1 at the SPring8 and BL6A at the Photon Factory. The experiments at the SPring8 were approved by the Japan Synchrotron Radiation Research Institute (JASRI), as proposal numbers 2004B0789, 2005A0877, 2005B0157, and 2006A1739. KM and DK were supported by Grants-in-Aid for Scientific Research in Priority Areas and the National Project on Protein Structural and Functional Analyses (Protein 3000) from the Ministry of Education, Culture, Sports, Science and Technology (MEXT) of Japan.

References

- Brunger AT, Adams PD, Clore GM, DeLano WL, Gros P, Grosse-Kunstleve RW, Jiang JS, Kuszewski J, Nilges M, Pannu NS, Read RJ, Rice LM, Simonson T, Warren GL (1998) Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr D* **54**: 905–921
- Chapados BR, Hosfield DJ, Han S, Qiu J, Yelent B, Shen B, Tainer JA (2004) Structural basis for FEN-1 substrate specificity and PCNA-mediated activation in DNA replication and repair. *Cell* **116**: 39–50
- Chen HW, North SH, Nakai H (2004) Properties of the PriA helicase domain and its role in binding PriA to specific DNA structures. *J Biol Chem* **279**: 38503–38512
- Cowtan K (1994) dm: An automated procedure for phase improvement by density modification. *Joint CCP4 and ESF-EACBM Newsletter on Protein Crystallography* **31**: 34–38
- de La Fortelle E, Bricogne G (1997) Maximum-likelihood heavy-atom parameter refinement for the multiple isomorphous replacement and multiwavelength anomalous diffraction methods. *Methods Enzymol* **276**: 472–494
- Heller RC, Marians KJ (2005a) The disposition of nascent strands at stalled replication forks dictates the pathway of replisome loading during restart. *Mol Cell* **17**: 733–743
- Heller RC, Marians KJ (2005b) Unwinding of the nascent lagging strand by Rep and PriA enables the direct restart of stalled replication forks. *J Biol Chem* **280**: 34143–34151
- Holm L, Sander C (1998) Touring protein fold space with Dali/FSSP. *Nucleic Acids Res* **26**: 316–319
- Jones JM, Nakai H (1997) The ϕ X174-type primosome promotes replisome assembly at the site of recombination in bacteriophage Mu transposition. *EMBO J* **16**: 6886–6895
- Jones TA, Zou JY, Cowan SW, Kjeldgaard M (1991) Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr A* **47** (Part 2): 110–119
- Kogoma T, Cadwell GW, Barnard KG, Asai T (1996) The DNA replication priming protein, PriA, is required for homologous recombination and double-strand break repair. *J Bacteriol* **178**: 1258–1264
- Kowalczykowski SC (2000) Initiation of genetic recombination and recombination-dependent replication. *Trends Biochem Sci* **25**: 156–165
- Lee EH, Kornberg A (1991) Replication deficiencies in priA mutants of *Escherichia coli* lacking the primosomal replication n' protein. *Proc Natl Acad Sci USA* **88**: 3029–3032
- Liu J, Marians KJ (1999) PriA-directed assembly of a primosome on D loop DNA. *J Biol Chem* **274**: 25033–25041
- Liu J, Xu L, Sandler SJ, Marians KJ (1999) Replication fork assembly at recombination intermediates is required for bacterial growth. *Proc Natl Acad Sci USA* **96**: 3552–3555
- Lorentzen E, Conti E (2005) Structural basis of 3' end RNA recognition and exoribonucleolytic cleavage by an exosome RNase PH core. *Mol Cell* **20**: 473–481
- Luscombe NM, Laskowski RA, Thornton JM (2001) Amino acid-base interactions: a three-dimensional analysis of protein-DNA interactions at an atomic level. *Nucleic Acids Res* **29**: 2860–2874
- Mahdi AA, Buckman C, Harris L, Lloyd RG (2006) Rep and PriA helicase activities prevent RecA from provoking unnecessary recombination during replication fork repair. *Genes Dev* **20**: 2135–2147
- Marians KJ (2000) PriA-directed replication fork restart in *Escherichia coli*. *Trends Biochem Sci* **25**: 185–189
- Masai H, Asai T, Kubota Y, Arai K, Kogoma T (1994) *Escherichia coli* PriA protein is essential for inducible and constitutive stable DNA replication. *EMBO J* **13**: 5338–5345
- McGlynn P, Al-Deib AA, Liu J, Marians KJ, Lloyd RG (1997) The DNA replication protein PriA and the recombination protein RecG bind D-loops. *J Mol Biol* **270**: 212–221
- Mizukoshi T, Tanaka T, Arai K, Kohda D, Masai H (2003) A critical role of the 3' terminus of nascent DNA chains in recognition of stalled replication forks. *J Biol Chem* **278**: 42234–42239
- Ng JY, Marians KJ (1996) The ordered assembly of the ϕ X174-type primosome. I. Isolation and identification of intermediate protein-DNA complexes. *J Biol Chem* **271**: 15642–15648
- North SH, Nakai H (2005) Host factors that promote transpososome disassembly and the PriA-PriC pathway for restart primosome assembly. *Mol Microbiol* **56**: 1601–1616
- Nurse P, Liu J, Marians KJ (1999) Two modes of PriA binding to DNA. *J Biol Chem* **274**: 25026–25032
- Otwinowski Z, Minor W (1997) Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol* **276**: 307–326
- Ouzounis CA, Blencowe BJ (1991) Bacterial DNA replication initiation factor priA is related to proteins belonging to the 'DEAD-box' family. *Nucleic Acids Res* **19**: 6953
- Perrakis A, Morris R, Lamzin VS (1999) Automated protein model building combined with iterative structure refinement. *Nat Struct Biol* **6**: 458–463
- Rigler R (1995) Fluorescence correlations, single molecule detection and large number screening. Applications in biotechnology. *J Biotechnol* **41**: 177–186
- Saito M, Go M, Shirai T (2006) An empirical approach for detecting nucleotide-binding sites on proteins. *Protein Eng Des Sel* **19**: 67–75
- Sasaki K, Ose T, Tanaka T, Mizukoshi T, Ishigaki T, Maenaka K, Masai H, Kohda D (2006) Crystallization and preliminary crystallographic analysis of the N-terminal domain of PriA from *Escherichia coli*. *Biochim Biophys Acta* **1764**: 157–160
- Shlomag J, Kornberg A (1980) An *Escherichia coli* replication protein that recognizes a unique sequence within a hairpin region in ϕ X174 DNA. *Proc Natl Acad Sci USA* **77**: 799–803
- Tanaka T, Masai H (2006) Stabilization of a stalled replication fork by concerted actions of two helicases. *J Biol Chem* **281**: 3484–3493
- Tanaka T, Mizukoshi T, Taniyama C, Kohda D, Arai K, Masai H (2002) DNA binding of PriA protein requires cooperation of the N-terminal D-loop/arrested-fork binding and C-terminal helicase domains. *J Biol Chem* **277**: 38062–38071
- Tanaka T, Taniyama C, Arai K, Masai H (2003) ATPase/helicase motif mutants of *Escherichia coli* PriA protein essential for recombination-dependent DNA replication. *Genes Cells* **8**: 251–261
- Terwilliger TC (2002) Automated structure solution, density modification and model building. *Acta Crystallogr D* **58**: 1937–1940
- Terwilliger TC, Berendzen J (1999) Automated MAD and MIR structure solution. *Acta Crystallogr D* **55**: 849–861
- Xu L, Marians KJ (2003) PriA mediates DNA replication pathway choice at recombination intermediates. *Mol Cell* **11**: 817–826
- Zavitz KH, Marians KJ (1992) ATPase-deficient mutants of the *Escherichia coli* DNA replication protein PriA are capable of catalyzing the assembly of active primosomes. *J Biol Chem* **267**: 6933–6940