Structural and biochemical analyses of hemimethylated DNA binding by the SeqA protein

Norie Fujikawa[1], Hitoshi Kurumizaka[1,2,3], Osamu Nureki[4], Yoshinori Tanaka[1,4], Mitsuyoshi Yamazoe[5], Sota Hiraga[5] and Shigeyuki Yokoyama[1,2,4,*]

[1]RIKEN Genomic Sciences Center, 1-7-22 Suehiro-cho, Tsurumi, Yokohama 230-0045, Japan, [2]Cellular Signaling Laboratory, RIKEN Harima Institute at SPring-8, 1-1-1 Kohto, Mikazuki-cho, Sayo, Hyogo 679-5148, Japan, [3]Waseda University School of Science and Engineering, 3-4-1 Okubo, Shinjuku-ku, Tokyo 169-8555, Japan, [4]Department of Biophysics and Biochemistry, Graduate School of Science, University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan and [5]Faculty of Medicine, Kyoto University, Konoe Yoshida, Sakyo-ku, Kyoto 606-8501, Japan

ABSTRACT

The *Escherichia coli* SeqA protein recognizes the 11 hemimethylated G-^{m}A-T-C sites in the *oriC* region of the chromosome, and prevents replication over-initiation within one cell cycle. The crystal structure of the SeqA C-terminal domain with hemimethylated DNA revealed the N^6-methyladenine recognition mechanism; however, the mechanism of discrimination between the hemimethylated and fully methylated states has remained elusive. In the present study, we performed mutational analyses of hemimethylated G-^{m}A-T-C sequences with the minimal DNA-binding domain of SeqA (SeqA$_{71-181}$), and found that SeqA$_{71-181}$ specifically binds to hemimethylated DNA containing a sequence with a mismatched ^{m}A:G base pair [G-^{m}A(:G)-T-C] as efficiently as the normal hemimethylated G-^{m}A(:T)-T-C sequence. We determined the crystal structures of SeqA$_{71-181}$ complexed with the mismatched and normal hemimethylated DNAs at 2.5 and 3.0 Å resolutions, respectively, and found that the mismatched ^{m}A:G base pair and the normal ^{m}A:T base pair are recognized by SeqA in a similar manner. Furthermore, in both crystal structures, an electron density is present near the unmethylated adenine, which is only methylated in the fully methylated state. This electron density, which may be due to a water molecule or a metal ion, can exist in the hemimethylated state, but not in the fully methylated state, because of steric clash with the additional methyl group.

INTRODUCTION

The *Escherichia coli* chromosome contains 19 130 G-A-T-C sites (1,2), which are normally methylated at the N6 position of the adenine bases (N^6-methyladenines) on both strands by the Dam methyltransferase (3). After replication, the adenine bases in the G-A-T-C sequences of the newly synthesized strand are immediately methylated, except for the 11 G-A-T-C sites within the replication origin, termed *oriC* (4). Actually, for nearly one-third of the cell cycle, the replicated duplexes in the *oriC* region remain in the hemimethylated state, where the N^6-methyladenine exists only in the parental strand. The next round of replication is not initiated from the hemimethylated *oriC* region, as the replication of chromosomal DNA is strictly regulated to occur once per cell cycle (synchronous replication initiation) (4). Therefore, the prolonged maintenance of the hemimethylated state at *oriC* is considered to prevent the premature initiation of replication (4).

The *E.coli* SeqA protein specifically binds to the hemimethylated G-^{m}A-T-C sequence in the *oriC* region, and prohibits the assembly of replication proteins, such as DnaA, onto the *oriC* region (5–9). Thus, SeqA prevents the premature initiation of chromosomal replication within one cell cycle, through its hemimethylated DNA binding. To do so, SeqA must discriminate the hemimethylated state from the fully methylated state. In fact, the affinity of SeqA for the hemimethylated G-^{m}A-T-C sequence is much stronger than that for the fully methylated sequence (10,11). The SeqA protein is composed of 181 amino acid residues (12), and has two functional domains (13,14). The N-terminal domain, consisting of amino acid residues 1–50, is responsible for the self-association of SeqA (13), while the minimal DNA-binding domain has been mapped on the C-terminal domain, consisting of amino acid residues 71–181 (SeqA$_{71-181}$) (14). Consistent with the results from the full-length protein (10,11), SeqA$_{71-181}$ exhibits stronger affinity for hemimethylated DNA than for fully methylated DNA (14).

The structure of the SeqA C-terminal segment (51–181 amino acids) has been determined in a complex with hemimethylated DNA (13). In the structure, SeqA specifically recognizes the N6-methyl group of methyladenine by two van der Waals contacts, and the propeller angle of the thymine base, which is the pairing partner with N^6-methyladenine, is significantly distorted by a base-specific hydrogen bond (13). In contrast, the discrimination mechanism between the

---

*To whom correspondence should be addressed. Tel: +81 45 503 9196; Fax: +81 45 503 9201; Email: yokoyama@biochem.s.u-tokyo.ac.jp

hemimethylated and fully methylated states was not eluci-dated, because fully methylated DNA can be accommodated without clashes with the C-terminal domain of SeqA in the modeled structure (13). Furthermore, the sequence-specific interactions between SeqA and the G-^{m}A-T-C sequence are limited to the 5′ three base pairs (G-^{m}A-T) (13), whereas previous biochemical studies showed that the entire G-^{m}A-T-C sequence is essential for the SeqA–DNA interaction (15).

In the present study, we first used gel-shift assays to analyze the binding of SeqA$_{71-181}$ to a 14mer DNA containing the hemimethylated G-^{m}A-T-C sequence with a series of muta-tions. These analyses revealed that all four of the base pairs are characteristically recognized by SeqA$_{71-181}$. We also unex-pectedly found that SeqA$_{71-181}$ can specifically bind DNA containing a hemimethylated sequence with a mismatched ^{m}A:G pair, G-^{m}A(:G)-T-C, as strongly as the normal hemimethylated G-^{m}A(:T)-T-C sequence. We determined two crystal structures of SeqA$_{71-181}$ complexed with a 10mer DNA, with either the G-^{m}A(:G)-T-C or G-^{m}A(:T)-T-C sequence. In the 2.5 Å structure with the mismatched hemimethylated sequence, the propeller angle of the guanine base, which is the pairing partner of the N^6-methyladenine, is distorted by about –35° to form a base-specific hydrogen bond with Asn152, similar to the thymine base in the structure with the normal hemimethylated DNA. The methyl group of the N^6-methyladenine is recognized by the van der Waals contacts of Thr151 and Asn152, as in the normal hemimethylated DNA (13). Furthermore, we found an electron density near the unmethylated adenine base, which is methylated only in the fully methylated state, in both crystal structures. The electron density, which is due to a water molecule or a metal ion, causes steric clash with the additional methyl group in the fully methylated state, and may be important in discriminating between hemimethylated and fully methylated DNA.

MATERIALS AND METHODS

Construction and purification of the SeqA$_{71-181}$ protein

The DNA fragment encoding SeqA$_{71-181}$, which contained amino acid residues 71–181, was ligated into the BamHI site of the pGEX-6P expression vector. SeqA$_{71-181}$ was expressed in the *E.coli* strain DH5α as an N-terminal GST fusion protein. The *E.coli* strain carrying the SeqA$_{71-181}$ expression vector was grown at 37°C, and isopropyl-β-D-thiogalactopyranoside (IPTG; 1 mM) was added at an OD$_{600}$ = 0.6 to induce the protein expression. After 3 h of cultivation, the cells were harvested, and the pellet was suspended in buffer A (50 mM Tris–HCl buffer pH 7.5, 500 mM NaCl, 10% glycerol, and 1 mM dithiothreitol). The cells were disrupted by sonication, and the sample was centrifuged at 27 000 *g* for 15 min at 4°C. The supernatant containing the GST-fused SeqA$_{71-181}$ protein was loaded onto a GSTrap column (5 ml) (Amersham Biosciences), and the column was washed with buffer A. The protein was eluted with buffer A containing 20 mM glutathione. The fractions containing the protein were dialyzed against buffer A, and PreScission protease (2 U per 100 μg of the fusion protein) (Amersham Biosciences) was added to uncouple the GST from the SeqA segment. The reaction proceeded at 4°C for 16 h, and the resulting GST was removed by GSTrap column chromatography. SeqA$_{71-181}$ was

recovered in the flow-through fraction, and was further purified on a Superdex 75 HR 10/30 column (Amersham Biosciences). Selenomethionine-labeled SeqA$_{71-181}$ (SeMet-SeqA$_{71-181}$) was expressed in the *E.coli* B834 strain, which was grown in the presence of L-selenomethionine.

DNA substrates

The HPLC-purified oligonucleotides for the DNA-binding assay (14mer; Fig. 1A) and for the crystallization (10mer; Fig. 2A) were purchased from Nihon Gene Research Laboratories. The nucleotide sequences are as follows: 10mer strand 1, 5′-AAG **GATC** CAA-3′; 10mer strand 2, 5′-TTG **GAGC** CTT-3′; 14mer strand 1, 5′-CT AAG **GATC** CAA GC-3′; 14mer strand 2, 5′-GC TTG **GATC** CTT AG-3′; and are derived from a SeqA-binding site in the *E.coli oriC* region. The G-A-T-C sequences that are recognized by SeqA are presented in bold, and the locations of the methyladenines are underlined. The hemimethylated duplex was prepared by annealing.

Assay for DNA binding

The hemimethylated double-stranded oligonucleotide (80 pmol) was incubated with SeqA$_{71-181}$ (200 pmol) at 37°C in 10 mM Tris–HCl buffer pH 7.5 containing 1 mM EDTA and 10% glycerol. The total volume of the reaction mixture was 20 μl. After a 30 min incubation, the samples were directly analyzed by 12% polyacrylamide gel electro-phoresis in 1× TBE buffer (90 mM Tris-borate and 2 mM EDTA) run at 3 V/cm for 60 min. The DNA bands were visualized by ethidium bromide staining.

Crystallization and data collection

The purified SeqA$_{71-181}$ protein was incubated with a hemimethylated 10mer double-stranded oligonucleotide (hemimethylated DNA) or a hemimethylated mismatched 10mer double-stranded oligonucleotide (hemimethylated T5*G DNA), which contains the ^{m}A:G base pair instead of the ^{m}A:T base pair at position 5. The resulting complexes were separated from the free protein and DNA by gel filtration chromatography on a HiLoad Superdex 75 column (Amersham Biosciences). The purified complexes were con-centrated up to 5 mg protein/ml, and co-crystals were obtained by the hanging drop method. The crystals of the SeqA$_{71-181}$–hemimethylated DNA complex were obtained in 90 mM HEPES buffer pH 7.5 containing 1.26 M tri-sodium citrate dihydrate and 10% glycerol, and belong to the hexagonal space group *P*622, with unit cell constants of *a* = 152.595 Å, *b* = 152.595 Å, *c* = 119.355 Å, α = β = 90°, γ = 120°. Two complexes were found in the asymmetric unit. A native diffraction data set was collected at 3.0 Å resolution from beamline BL41XU at SPring-8 in Harima. The data were processed and were scaled using the DENZO and SCALEPACK programs (16). The crystals of SeqA$_{71-181}$ complexed with hemimethylated T5*G DNA were obtained in 100 mM sodium cacodylate buffer pH 6.5 containing 200 mM calcium acetate hydrate and 18% polyethylene glycol 8000, and belong to the orthorhombic space group *P*2$_1$2$_1$2$_1$, with unit cell constants of *a* = 33.432 Å, *b* = 65.061 Å, *c* = 77.176 Å. One complex was found in the asymmetric unit. A native diffraction data set was collected at 2.5 Å resolution from beamline BL41XU at SPring-8 in Harima.
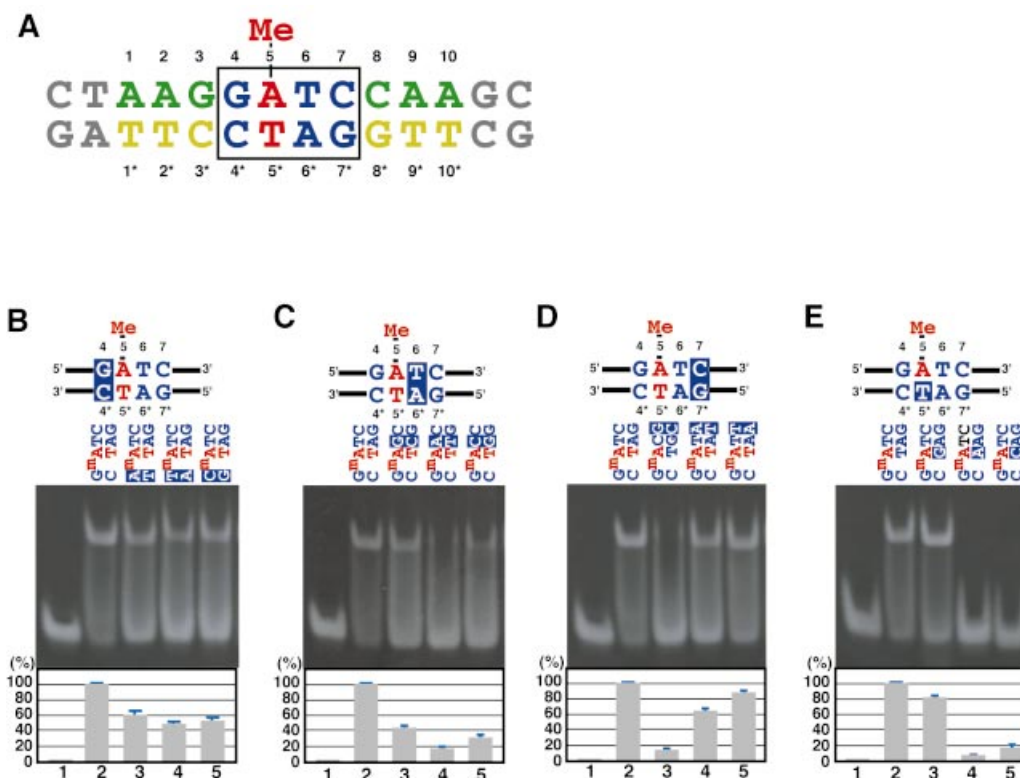
Figure 1. SeqA$_{71-181}$–DNA binding analyses. (**A**) The hemimethylated G-^{m}A-T-C 14mer DNA used in the DNA binding analysis. The region corresponding to the hemimethylated G-^{m}A-T-C 10mer DNA used in the crystallization of the SeqA$_{71-181}$–DNA complex is colored, and is numbered. The recognition sequence of SeqA is colored blue in the box, and the hemimethylated ^{m}A:T base pair is colored red. 'Me' indicates the N6-methyl group. Numbers without asterisks indicate base positions from the 5′ end of the methylated strand, and numbers with asterisks indicate base positions from the 3′ end of the unmethylated strand. (**B–D**) Effect of base pair replacements at the G4:C4* (**B**), T6:A6* (**C**) and C7:G7* (**D**) base pairs. Replaced base pairs are indicated by white letters on the top of each panel. Lane 1 is a negative control experiment without protein, and lane 2 is a positive control. (**E**) Effect of mismatch replacements at the ^{m}A5:T5* position on the DNA binding. Lane 1 is a negative control experiment without protein, and lane 2 is a positive control. The T5* residue was replaced by G (lane 3), A (lane 4) and C (lane 5). Graphic representations of the DNA binding of the SeqA$_{71-181}$ mutants are presented in the bottom panels (B–E). The amounts of complex formation relative to that accomplished by the wild-type SeqA$_{71-181}$ protein are presented. Average and SD values from three independent experiments are presented.

Structure determination and refinement

The crystal structure of the SeqA-hemimethylated DNA was solved by the multiwavelength anomalous dispersion (MAD) method (17), using the selenomethionine-labeled SeqA$_{71-181}$. The scaled data set at the peak wavelength of the MAD experiment up to 3.2 Å was used to calculate the normalized structure factor with DREAR (18), and the resulting data were input into SnB (18) to locate the selenium atoms. After trials with SnB, seven consistent peaks were picked out of the 12 atoms expected in the asymmetric unit, and were input into program SHARP (19) to calculate the initial phases and the heavy atom refinement. The resulting initial phases were refined with density modification using SOLOMON (20). The map was improved by density modification procedures, such as solvent flattening and non-crystallographic symmetry (NCS) averaging, with the program DM (20), which extended the phase up to 3.0 Å using the native data. An atomic model was fitted into the electron density map using the graphics program O (21), and was refined by energy minimization and simulated annealing procedures with CNS (22). The final model contains 1816 atoms of SeqA$_{71-181}$, 808 atoms of DNA and 131 water molecules. The Ramachandran plot of the final structure showed 85% of the residues in the most favorable regions and 15% of the residues in the additionally allowed regions. The structure of the orthorhombic crystal form of the complex with hemimethylated T5*G DNA was solved by molecular replacement with the program AMORE (20), using the crystal structure of the hexagonal form as a search model. The refined model has an R_{work} of 23.1% and an R_{free} of 28.8% at 2.5 Å resolution. The final model contains 1816 atoms of SeqA$_{71-181}$, 808 atoms of DNA and 95 water molecules. The Ramachandran plot of the final structure showed 90.3 % of the residues in the most favorable regions, 8.7% of the residues in the additionally allowed regions, 1.0% of the residues in the generously allowed regions. Graphic figures were created using the RIBBONS (23). The atomic coordinates of SeqA$_{71-181}$ complexed with the hemimethylated 10mer have been deposited (RCSB id code: rcsb005276, PDB id code: 1IU3), as have those with the hemimethylated mismatched 10mer (RCSB id code: rcsb005580, PDB id code: 1J3E).

RESULTS AND DISCUSSION

Mutational analysis of the hemimethylated G-^{m}A-T-C sequence in DNA binding by SeqA$_{71-181}$

According to the structural analysis by Guarne *et al.*, three base pairs, G:C-^{m}A:T-T:A, within the hemimethylated G-^{m}A-T-C sequence were recognized by the base-specific
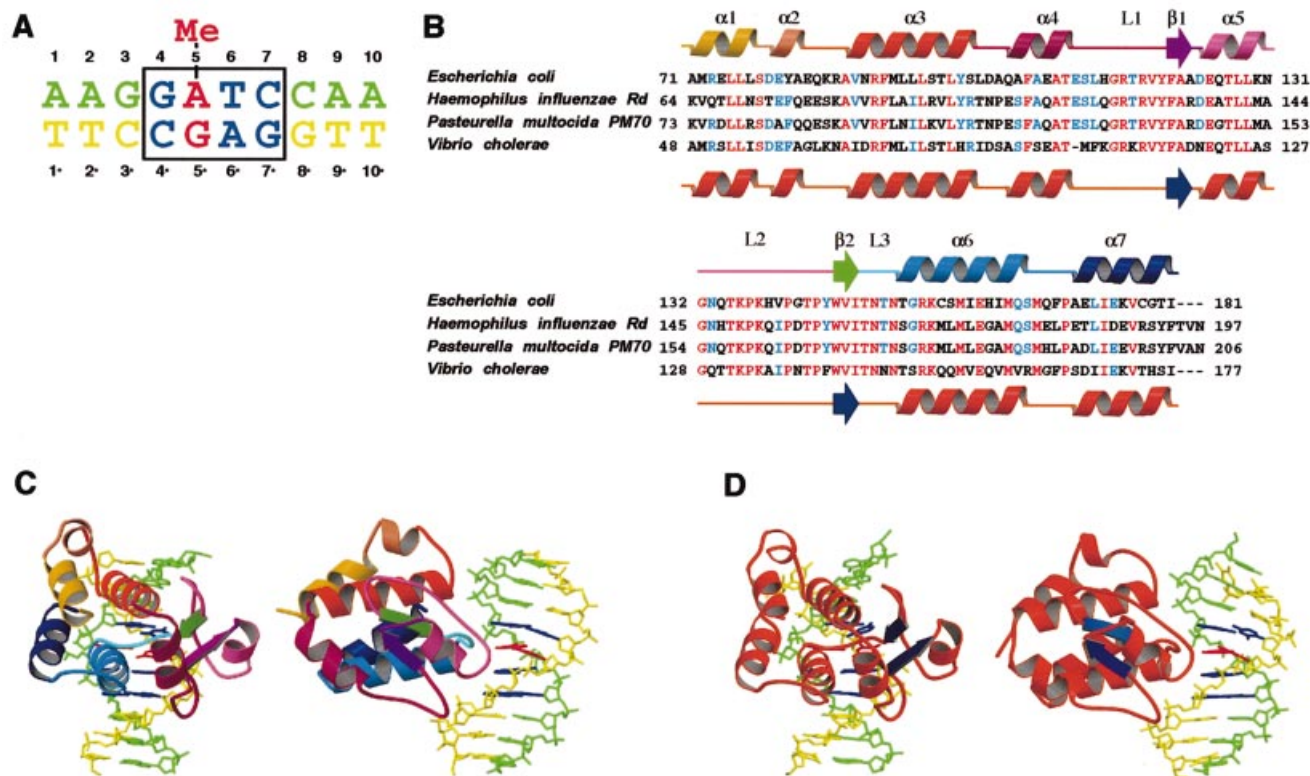
Figure 2. Crystal structures of SeqA$_{71-181}$ complexed with a T5*G mismatch hemimethylated DNA and a normal hemimethylated DNA. (**A**) The T5*G mismatch hemimethylated G-^{m}A-T-C DNA used in the structural analysis of the SeqA$_{71-181}$–DNA complex. The recognition sequence of SeqA is colored blue in the box, and the hemimethylated ^{m}A:T base pair is colored red. 'Me' indicates the N6-methyl group. Numbers without asterisks indicate base positions from the 5′ end of the methylated strand, and numbers with asterisks indicate base positions from the 3′ end of the unmethylated strand. (**B**) Secondary structure of SeqA$_{71-181}$. The amino acid sequences from the *Haemophilus influenzae* Rd, *Pasteurella multocida* PM70 and *Vibrio cholerae* SeqA proteins, corresponding to the *E.coli* SeqA$_{71-181}$ protein, are aligned. The α-helix and β-sheet regions are indicated by yellow and green boxes, respectively. Conserved and semi-conserved amino acid residues are indicated by red and blue letters, respectively. The secondary structure of SeqA$_{71-181}$ complexed with the T5*G mismatch hemimethylated DNA is presented in the top row, and that with the authentic hemimethylated DNA is presented in the bottom row. (**C**) Overall structures of the SeqA$_{71-181}$–T5*G hemimethylated DNA complex. Top and side views are shown. The DNA strand containing the N6-methyladenine is shown in green, and its complementary strand is shown in yellow. The recognition bases are colored blue, and the hemimethylated ^{m}A:T base pairs are shown in red. The colors of the protein correspond to those in (B). (**D**) Overall structures of the SeqA$_{71-181}$–hemimethylated DNA complex. Top and side views are shown.

interactions with the C-terminal domain of SeqA (13). To test the importance of these base-specific interactions in SeqA$_{71-181}$ binding, we performed a comprehensive mutational study on a 14mer DNA with the hemimethylated G-^{m}A-T-C sequence (Fig. 1A). First, we studied the G4:C4* base pair, in which the G4 base interacts with Asn150 in the crystal structure (13). As shown in Figure 1B, replacements of the G4:C4* base pair by A:T, T:A and C:G moderately reduced the DNA binding ability, although detectable amounts (40–60% of that of the wild-type sequence) of the complexes were still formed. Therefore, the G4:C4* base pair contributes somewhat to the specific binding. Next, the T6:A6* base pair, which is methylated in the fully methylated state but not in the hemimethylated state, was examined. In the crystal structure, the T6 base hydrogen-bonds with Asn152 (13). When the T6:A6* base pair was replaced by A6:T6*, the DNA binding was significantly decreased (Fig. 1C, lane 4). The replacements of the T6:A6* base pair by G6:C6* and C6:G6* also decreased the SeqA binding. Thus, the T6:A6* base pair contributes to the specific binding, as indicated by the crystal structure (13). Interestingly, the C7:G7* base pair has no

base-specific interactions in the crystal structure (13). Consistent with this observation, the replacements of the C7:G7* base pair by A:T and T:A did not cause a significant decrease in the SeqA$_{71-181}$ DNA binding (Fig. 1D, lanes 4 and 5). However, the replacement by the G:C base pair significantly reduced the SeqA$_{71-181}$ DNA binding (Fig. 1D, lane 3). Therefore, the C7:G7* base pair functions to eliminate sequences containing G:C at this position, in some indirect way. Further analyses will be required to elucidate how this base pair replacement affects the hemimethylated DNA binding by SeqA.

In the crystal structure, the propeller angle of the thymine base, which is the pairing partner of the N^6-methyladenine, is significantly distorted by a base-specific hydrogen bond between the O4 of T5* and the side chain of Asn150 (13). In the present study, the T5* base of the A5:T5* base pair, but not A5, was replaced by adenine or cytosine (A5:A5* and A5:C5*, respectively), so that the hydrogen bond with Asn150 was disrupted. These mismatched mutations abolished the SeqA$_{71-181}$ binding (Fig. 1E, lanes 4 and 5). Thus, the T5* residue is crucial for the specific DNA binding. Surprisingly,

the replacement of A5:T5* with A5:G5* did not affect the specific DNA binding (Fig. 1E, lane 3). This implied that Asn150 formed a hydrogen bond with the O6 of G5*, in place of the O4 of T5*.

Structure of the SeqA$_{71-181}$-hemimethylated T5*G DNA complex

To study the molecular mechanism of SeqA binding to hemimethylated DNA with the mismatched A5:G5* base pair, we determined the crystal structure of SeqA$_{71-181}$ complexed with a 10mer hemimethylated T5*G DNA containing the G-^{m}A(:G)-T-C sequence (the SeqA$_{71-181}$·T5*G complex; Fig. 2A) at 2.5 Å resolution (Fig. 2). The 10mer is the minimal size for SeqA$_{71-181}$ binding of G-^{m}A-T-C DNA (14). The crystals of the SeqA$_{71-181}$·T5*G complex belong to the orthorhombic space group $P2_12_12_1$ (Table 1). We also obtained crystals of SeqA$_{71-181}$ bound with the normal hemimethylated 10mer DNA containing the G-^{m}A(:T)-T-C sequence (the SeqA$_{71-181}$·10mer complex; Fig. 1A), which belong to the hexagonal space group $P622$ (Table 1). On the other hand, the crystals with the hemimethylated 12mer, obtained by Guarne *et al.* [the SeqA$_{51-181}$·12mer complex (13)], belong to either the $P2_12_12$ or $P4_2$ space group. It is interesting that the space groups of these SeqA–DNA crystals differ from each other. Nevertheless, the SeqA$_{71-181}$ structure in the SeqA$_{71-181}$·T5*G complex is essentially the same as that in the SeqA$_{71-181}$·10mer complex (r.m.s.d. 0.6 Å; Fig. 2B–D) and the SeqA$_{51-181}$·12mer complex (r.m.s.d. 0.7 Å).

In the SeqA$_{71-181}$·T5*G complex, the electron density of the N6-methyl group of A5 is clearly observed, and the A5 base-pairs with the G5* base through a hydrogen bond (Fig. 3A). The N6-methyl group of A5 is specifically recognized by two van der Waals contacts, with the main chain CO group of Thr151 and the main chain NH group of Asn152, which are the same as those of the normal hemimethylated DNA [Fig. 3; Guarne *et al.* (13)]. The distances from the main chain CO group of Thr151 and the main chain NH group of Asn152 to the methyl group of A5 are about 3.5 Å, thus allowing strong van der Waals contacts to form between them (Fig. 3A and B).

Propeller twist of the guanine base that pairs with N^6-methyladenine

In the present structure of the SeqA$_{71-181}$·10mer complex, the T5* base, which is the pairing partner of the methyl-A5, exhibits a significantly different propeller angle (about –35°), which is about three times larger than that of the protein-free hemimethylated A:T base pair (about –10°) [Baikalov *et al.* (24); Fig. 4A, C and E]. This distortion of the thymine base was also observed in the SeqA$_{51-181}$·12mer complex (13). In the SeqA$_{71-181}$·10mer complex, the T5* distortion places the O4 atom of T5* in close enough proximity to form a hydrogen bond with the δNH$_2$ group of Asn150 (Fig. 3C), indicating that the distortion is essential for the base-specific recognition by SeqA. The propeller angle of the mismatched G5* base in the hemimethylated T5*G DNA is also distorted in the SeqA$_{71-181}$·T5*G complex, like the case of T5* in the SeqA$_{71-181}$·10mer complex (Fig. 4B and D). This G5* distortion actually allows the O6 of G5* to form a hydrogen bond with Asn150 (Figs 3A and 4D). Therefore, these results indicate that the base distortion and the hydrogen bonding with the

base, which pairs with the methyladenine, are essential for the hemimethylated G-^{m}A-T-C recognition by SeqA.

Recognition of the unmethylated T6:A6* base pair, which is methylated only in the fully methylated state

Previous structural analyses revealed the mechanism for the N6-methyl group recognition by SeqA (13), but the mechanism of SeqA discrimination between hemimethylated DNA and fully methylated DNA has not been determined. In the present study, we found an electron density near A6*, which is methylated only in the fully methylated state (Fig. 5A). This electron density near A6* was not observed in the previous structural analysis with the SeqA$_{51-181}$·12mer complex (13). This electron density could be due to either a water molecule or a metal ion, such as sodium, since sodium was used in the crystallization. Although we cannot exclude the possibility that the electron density is due to a sodium ion, we prefer the explanation that it is due to a water molecule, for the following reasons. (i) The electron density is located in a position where it can form hydrogen bonds with the main chain CO group of Asn150 and the N7 of A6* with an angle of 102°, which is close to the favored angle for these bonds (Figs 3C and 5A). (ii) When the ^{m}A5:T5* and T6:A6* base pairs of the SeqA$_{71-181}$·10mer complex (Fig. 5B, red) were superimposed on those of the SeqA$_{51-181}$·12mer complex (Fig. 5B, blue), the ^{m}A5:T5* base pair overlapped well. However, the A6* base, which may directly interact with the atom in the electron density, did not overlap with the corresponding base of the SeqA$_{51-181}$·12mer complex (Fig. 5B). This difference in the A6* base may be due to the water-mediated hydrogen bonding, which is found in the SeqA$_{71-181}$·10mer complex but not in the SeqA$_{51-181}$·12mer complex. (iii) The *B*-factor of this atom is low (19.78 Å^2), indicating its stable binding to the SeqA·DNA complex. These results support the idea that the electron density is due to a water molecule. However, the lengths of these water-mediated hydrogen bonds are shorter (2.4 and 2.5 Å) than a regular hydrogen bond (about 2.8 Å) (Fig. 5A and C). The strong binding of the water molecule to the SeqA·DNA complex may shorten the lengths of the hydrogen bonds.

If the electron density is due to a salt ion, then the hemimethylated DNA binding of SeqA could be affected by the salt concentration and species. Therefore, we next tested the hemimethylated DNA binding of SeqA$_{71-181}$ with different salt concentrations (Fig. 6A and B) and species (Fig. 6C–G). SeqA$_{71-181}$ specifically bound to the hemimethylated DNA, but not to the unmethylated or fully methylated DNAs, in the presence of 300 mM NaCl (Fig. 6A). The specific SeqA$_{71-181}$ binding to hemimethylated DNA was not affected when the NaCl concentration was reduced to 30 mM (Fig. 6B). In addition, SeqA$_{71-181}$ specifically bound to the hemimethylated DNA when NaCl was replaced by KCl, Li$_2$SO$_4$, MgSO$_4$ or MgCl$_2$ (Fig. 6C–F). These results confirm that salt ions do not affect the hemimethylated DNA binding of SeqA. Therefore, the electron density that interacts with the main chain CO group of Asn150 and the N7 of A6* may not be a salt ion, but a water molecule.

If the electron density near Asn150 and A6* is due to a water molecule, then it is bound in the major groove side of the T6:A6* base pair, through two hydrogen bonds with the main chain CO group of Asn150 and the N7 of A6* (Figs 3C and

Table 1. X-ray data collection, phasing and refinement statistics

	SeMet crystal Peak	Edge	Remote1	Remote2	Native	T5*G mutant
Data collection						
Beamline	BL41XU, SPring-8					
Wavelength (Å)	0.9798	0.9800	0.9839	0.9742	0.9797	0.9740
Resolution (Å)	50–3.2	50–3.2	50–3.2	50–3.2	50–3.0	50–2.5
(last shell)	(3.26–3.2)	(3.26–3.2)	(3.26–3.2)	(3.26–3.2)	(3.05–3.0)	(2.54–2.5)
Unique reflections	13 201	13 112	12 864	12 890	16 548	6126
Redundancy	11.9	11.7	11.1	11.3	11.6	5.5
Completeness (%)						
All data	94.6	94.1	93.0	93.4	98.5	97.3
Last shell	93.8	93.3	89.7	90.9	95.9	91.5
$I/\sigma(I)$						
All data	8.2	7.8	7.3	7.1	9.7	12.0
Last shell	2.9	2.7	2.8	2.8	2.2	4.9
R_{sym}[a] (%)						
All data	9.3	8.2	8.6	9.5	7.2	6.1
Last shell	40.2	41.9	39.3	40.2	40.0	21.2
Phasing statistics						
No. of sites	7	7	7	7		
Phasing power						
Iso	0.31	–	0.53	0.54		
	(0.20)[b]		(0.37)[b]	(0.51)[b]		
R_{cullis}[c]						
Iso	0.98	–	0.94	0.94		
	(0.97)[b]		(0.91)[b]	(0.91)[b]		
Ano	0.82	0.93	0.99	0.89		
FOM			0.36	(0.32)[b]		
Refinement statistics						
Resolution (Å)					3.0	2.5
Protein atoms					1816	1816
DNA atoms					808	808
Water oxygens					131	95
R_{work} (%)					23.8	22.6
R_{free}[d] (%)					28.6	29.4
Average B-factor (Å^2)					79.0	34.2
Minimum B-factor (Å^2)					27.6	7.3
Maximum B-factor (Å^2)					160.6	108.0
R.m.s.d. bond length (Å)					0.0320	0.00656
R.m.s.d. bond angles (°)					3.16	1.19
R.m.s.d. dihedrals (°)					22.8	19.3
R.m.s.d. impropers (°)					2.67	1.16

[a]$R_{sym} = \Sigma |I_{avg} - Ii|/\Sigma Ii$.
[b]Values in parentheses represent those for centric reflections.
[c]$R_{cullis} = \Sigma \|F_{PH} + F_P| - F_H(calc)|/\Sigma |F_{PH}|$.
[d]R_{free} is calculated for 10% of randomly selected reflections excluded from refinement.

5A). This T6:A6* base pair is methylated in the fully methylated state, but not in the hemimethylated state. Interestingly, when a methyl group is modeled onto the N7 position of A6*, the van der Waals radius of the A6* methyl group overlaps with that of the water molecule (Fig. 5B and C). Thus, the water molecule would be excluded due to steric hindrance with the additional methyl group of A6* in the fully methylated state, and the water-mediated hydrogen bonds would be destroyed. Therefore, we suggest that the water-mediated hydrogen bonds, which can be formed only in the hemimethylated state (Fig. 5B), may be the key mechanism for the discrimination against the fully methylated state (Fig. 5C).

The L3 loop

The hemimethylated ^{m}A:T pair of the DNA associates with the L3 loop, which is composed of five conserved amino acid residues (Ile148–Thr149–Asn150–Thr151–Asn152, Fig. 2B). The L3 loop is characteristically stretched between β2 and α6,

and lies co-planarly with the hemimethylated base pair on the major groove side in both the SeqA$_{71-181}$·T5*G and SeqA$_{71-181}$·10mer complexes (Fig. 2C and D). The N6-methyl group of A5 is specifically recognized by two van der Waals contacts with the main chain CO group of Thr151 and the main chain NH group of Asn152; both residues are located in the L3 loop (Figs 2B and 3). As expected, the T151A mutant (Thr151 to Ala) retained the ability to recognize the hemimethylated DNA (Fig. 7, lane 15), indicating that the main chain atoms, but not the side chain atoms, of Thr151 are actually involved in the recognition of the N6-methyl group of A5 in the hemimethylated DNA. However, as shown in Figure 7 (lane 15), the DNA binding ability of the T151A mutant was reduced to about 40% of that of the wild-type SeqA$_{71-181}$ protein. Thr151 formed two intramolecular hydrogen bonds, with Thr149 and Lys156 (data not shown). The alanine mutations of Thr149 (T149A) and Lys156 (K156A) also caused significant deficiencies in the DNA binding (Fig. 7, lanes 10 and 21), indicating that the defective DNA binding of
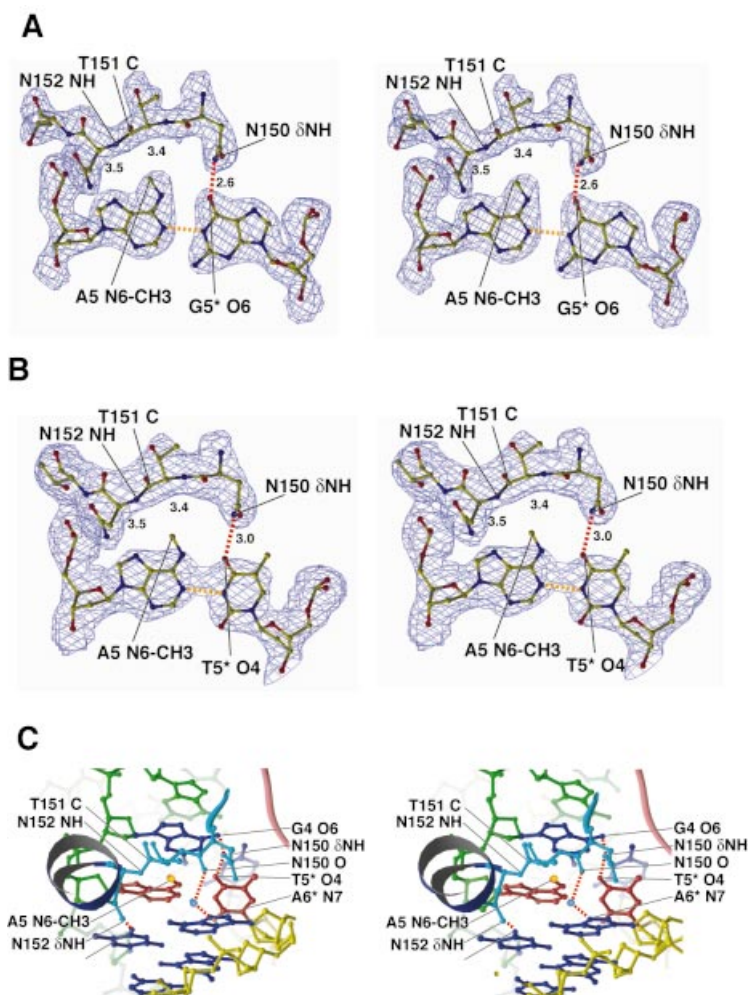
Figure 3. The SeqA–DNA interactions. (**A**) Stereo view of the $(2|F_o| - |F_c|)$ electron density map for the L3 loop and the $^mA5:G5^*$ base pair. The hydrogen bond between the base pair is represented by the yellow dashed line, and the hydrogen bond between Asn150 and G5* is indicated by the red dashed line. The number on the red dashed line indicates the length (Å) of the hydrogen bond. The purple arrows indicate van der Waals interactions, and the numbers on the purple arrows indicate the distances (Å) between the main chain CO group of Thr151 or the NH group of Asn152 and the methyl group. (**B**) Stereo view of the $(2|F_o| - |F_c|)$ electron density map for the L3 loop and the $^mA5:T5^*$ base pair. (**C**) Stereo view of the interactions between the L3 loop and the major groove of the hemimethylated $^mA:T$ base pair. Colors correspond to those in Figure 2C and D. The hemimethylated $^mA:T$ base pair is colored red, and the N6-methyl group of A5 is presented in a gold sphere. Red dashed lines indicate hydrogen bonds, and purple arrows indicate van der Waals interactions with the N6-methyl group of A5. The water molecule bound by the hydrogen bonds is presented in a blue sphere.

the T149A and K156A mutants could be due to a perturbation of the intramolecular interaction with Thr151. These intramolecular interactions maintain the stretched structure of the L3 loop, and may be important for precisely placing the L3 loop in the major groove to form the van der Waals contacts with the methyl group of A5.

Amino acid residues involved in the base-specific recognition

In addition to hydrogen bonding with the O4 of T5*, the δNH_2 group of Asn150 also hydrogen-bonds with the O6 of G4 (Fig. 8A). When Asn150 was replaced by alanine (N150A) or aspartic acid (N150D), which should disrupt both hydrogen bonds with T5* and G4, the hemimethylated DNA binding ability was significantly decreased (Fig. 7, lanes 11 and 12). Consistent with the observations in the crystal structures, these results indicate that the δNH_2 group of Asn150 significantly contributes to the base-specific DNA binding by SeqA.

Interestingly, the mutations of Asn150 to lysine (N150K) and glutamine (N150Q) also caused deficiencies in the hemimethylated DNA binding, but ~40% of the complexes were formed, as compared with that of the wild-type SeqA$_{71-181}$ protein (Fig. 7, lanes 13 and 14). The lysine and glutamine mutations may retain some of the hydrogen bonds formed by Asn150.

In the T6:A6* base pair, the O4 of T6 forms a hydrogen bond with the δNH_2 group of Asn152 (Figs 3C and 8A). When Asn152 was replaced by aspartic acid (N152D), the SeqA$_{71-181}$ DNA binding ability was almost completely abolished (Fig. 7, lane 17). The mutation of Asn152 to alanine (N152A) also caused a significant reduction in the DNA binding (~25% of that of the wild-type SeqA$_{71-181}$ protein) (Fig. 7, lane 16). Therefore, the hydrogen bond between T6 and the Asn152 side chain is important in the specific DNA binding by SeqA. However, the N152K and N152Q mutants (Asn152 to lysine and glutamine, respectively) retain ~75 and 50% (Fig. 7, lanes
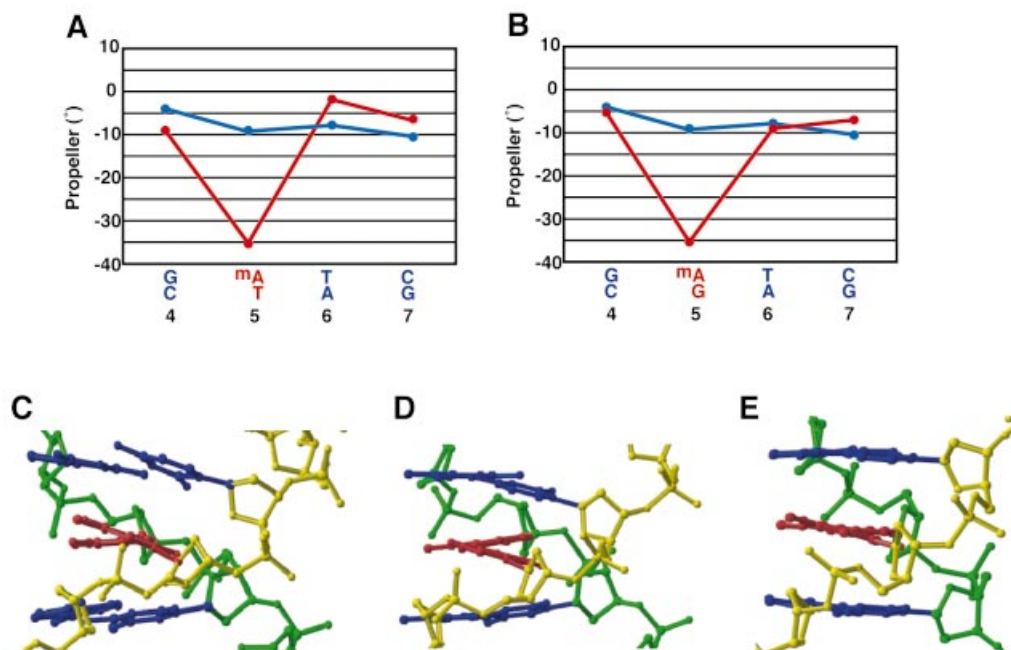
Figure 4. The DNA structure in the SeqA$_{71-181}$–DNA complex. (**A**) Graphic representation of the base propeller angles of the authentic hemimethylated DNA, calculated by CURVES (31). (**B**) Graphic representation of the base propeller angles of the T5*G hemimethylated DNA, calculated by CURVES (31). The vertical axis indicates the propeller angle of each base pair, and the horizontal axis indicates the base number. The red and blue lines indicate the propeller angles of the SeqA$_{71-181}$-bound hemimethylated DNA and the protein-free hemimethylated DNA (24), respectively. (**C–E**) Structural comparison of the SeqA$_{71-181}$-bound hemimethylated DNA (**C**), the SeqA$_{71-181}$-bound T5*G hemimethylated DNA (**D**) and the protein-free hemimethylated DNA (**E**). The recognition bases are colored blue, and the methylated and unmethylated strands are presented in green and yellow, respectively.

18 and 19), respectively, of the hemimethylated DNA binding abilities, and the complex with the N152K mutant showed unusually slower migration (lane 18). These polar residues may mimic the hydrogen bond formed by Asn152.

Interactions between SeqA$_{71-181}$ and the 5′-phosphate groups of G8*, G7* and A6*

The SeqA$_{71-181}$-bound DNA molecule is globally distorted; the minor groove is significantly narrower than that of B-form DNA (Fig. 8B). The 5′-phosphate groups of G8*, G7* and A6* in the unmethylated strand strongly interact with the amino acid residues from the L1, L2 and α6 regions (Fig. 8A). To test the importance of these interactions, we performed alanine scanning mutagenesis in the L1, L2 and α6 regions (Fig. 7). The alanine mutations of Arg116 and Arg118 (R116A and R118A, respectively) in the L1 loop significantly reduced the hemimethylated DNA binding (Fig. 7, lanes 3 and 5), whereas the alanine mutation of Thr117 (T117A) only slightly affected the DNA binding (lane 4). Consistently, in the present SeqA$_{71-181}$·10mer complex structure, the main chain NH group of Arg116 interacts with the G7* phosphate group, and the aliphatic part of the Arg116 side chain forms a van der Waals contact with the deoxyribose moiety of G8* (Fig. 8A). Furthermore, the εN of Arg118 electrostatically interacts with the A6* phosphate group, in addition to its interactions with the main chain NH group and the G7* phosphate group (Fig. 8A). In α6, Arg155, which interacts with the G8* phosphate group (Fig. 8A), also contributes to the DNA binding, because the R155A mutant (Arg155 to alanine) was significantly defective in this function (Fig. 7, lane 20). These

essential interactions with the backbone phosphate groups of the unmethylated strand may be important for the distorted DNA structure found in the SeqA$_{71-181}$·DNA complex (Fig. 8B).

As shown in Figure 7, the N133A, Q134A, T135A and K136A mutants in the L2 loop each caused either no decrease or only a slight decrease in the DNA binding ability (lanes 6–9), although the side chain NH$_2$ group of Arg133 interacts with the phosphate group of A6* in the SeqA$_{71-181}$·10mer structure (Fig. 8A). Therefore, in contrast to the L1 loop, the amino acid residues in the L2 loop do not seem to be essential for the DNA binding by SeqA. Thus, these mutational analyses of SeqA$_{71-181}$ have confirmed the sequence- and hemimethyl-specific DNA recognition mechanisms indicated by the crystal structures.

Comparison with other methyl-DNA-binding proteins

The crystal structure of another hemimethylated G-^{m}A-T-C-binding protein, the *E.coli* MutH protein, has been reported (25). The MutH protein is an essential protein for methyl-directed DNA mismatch repair, to correct errors made during DNA replication, and cleaves only the unmethylated strand in the hemimethylated G-^{m}A-T-C duplex. MutH exhibits structural homology to the PvuII and EcoRV type II endonucleases (25). The mechanism of the hemimethylated G-^{m}A-T-C recognition by MutH is not yet understood, but it probably differs from that of SeqA, because the structure of MutH is totally different from that of the SeqA DNA-binding domain.

The crystal structures of other hemimethylated DNA-binding proteins, M.HhaI, M.HaeIII and M.TaqI, have been
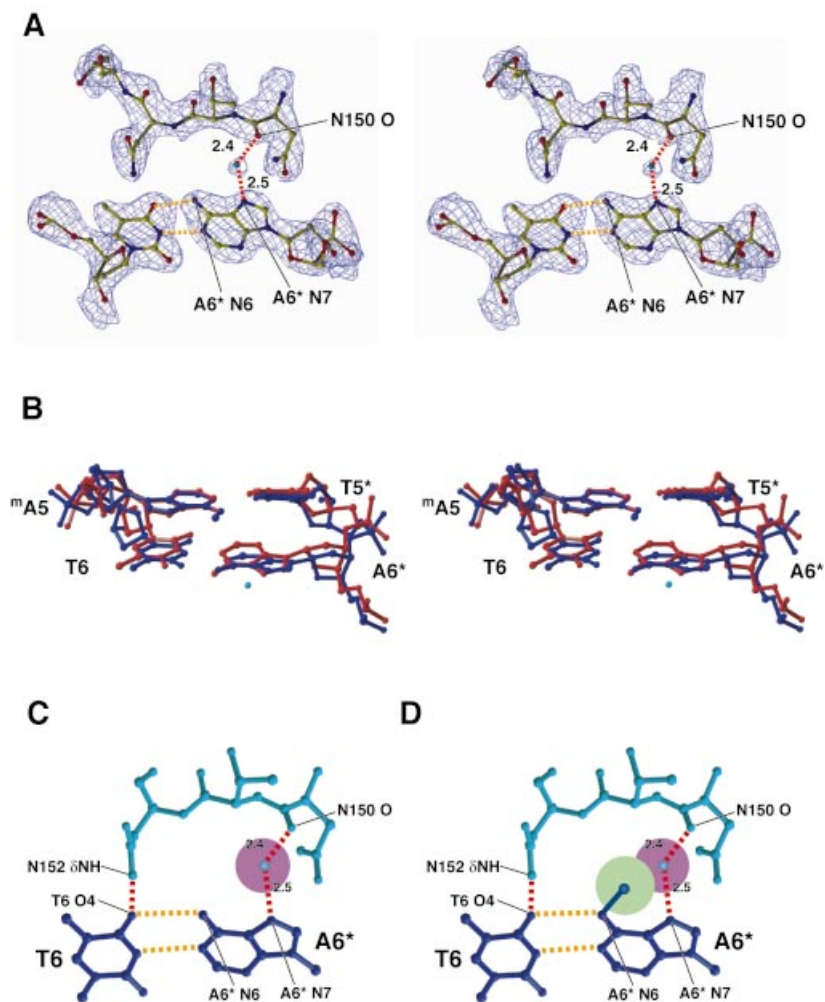
Figure 5. Discrimination of the hemimethylated state from the fully methylated and unmethylated states. (**A**) Stereo view of the $(2|F_o| - |F_c|)$ electron density map for the L3 loop and the T6:A6* base pair. The hydrogen bonds between the base pair are represented by yellow dashed lines, and the water-mediated hydrogen bonds between Asn150 and A6* are indicated by red dashed lines. The water molecule is presented in a blue sphere. The numbers on the red dashed lines indicate the lengths (Å) of the hydrogen bonds. (**B**) Stereo view of the hemimethylated A:T and non-methylated T:A base pairs complexed with SeqA. The $^mA5:T5^*$ and T6:A6* base pairs of the SeqA$_{71–181}$·10mer complex (red) were superimposed on those of the SeqA$_{51–181}$·12mer complex (13; blue). (**C**) The interaction between the L3 loop and the T6:A6* base pair found in the present complex structure. (**D**) The N6-methyl group is modeled at A6*, which is methylated in the fully methylated state. The van der Waals radii of the oxygen (1.4 Å) of the water molecule and the methyl group (2.0 Å) are indicated in purple and green circles, respectively.

reported in the complex forms with their target DNA molecules (26–28). These proteins, which have a common fold, are *S*-adenosyl-L-methionine-dependent methyltransferases (AdoMet-dependent MTases), and methylate the C5 position of a cytosine (M.HhaI and M.HaeIII) or the N6 position of an adenine (M.TaqI) in their recognition sequences. These MTases, which are components of the bacterial restriction–modification systems, sequence specifically bind to their target DNA, and flip out the cytosine or adenine base from the DNA helix. The flipped cytosine or adenine is then methylated at the active center of the protein (28,29). Therefore, the methyl recognition mechanism of MTase differs from that of SeqA.

In addition to these bacterial methylated DNA-binding proteins, the solution structure of the methyl-binding domain of the human MBD1 protein complexed with its target, fully methylated DNA, has been reported (30). MBD1 recognizes the cytosine methyl groups in both strands in an asymmetric manner, using a large hydrophobic patch formed by the side chains of five amino acid residues. In contrast, SeqA recognizes the adenine methyl group, with a small van der Waals surface made up of only two main chain atoms. The small methyl-binding surface of SeqA may be suitable for recognizing only one methyl group, but not two methyl groups in both strands.

In the present study, we have determined the molecular mechanism of the hemimethylated DNA recognition by SeqA, and have suggested a mechanism for its exclusion of fully methylated DNA. The unique structure of the hemimethylated G-^{m}A-T-C sequence and the mechanism of its recognition by SeqA provide important insights into how SeqA specifically recognizes the hemimethylated G-^{m}A-T-C sequences among
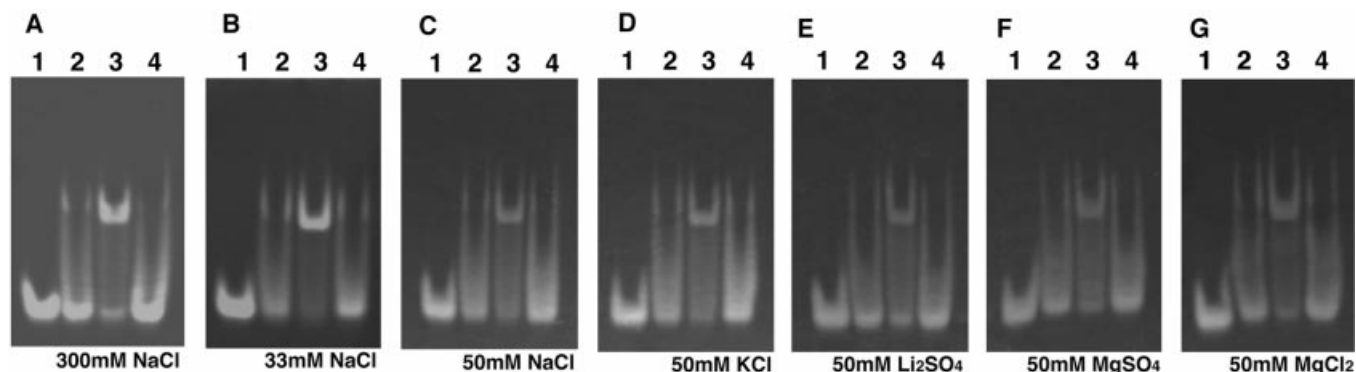
Figure 6. Effects of salt concentration and species on SeqA$_{71-181}$–DNA binding. Double-stranded 14mer oligonucleotides were used in the DNA binding assay. The unmethylated (lane 2), hemimethylated (lane 3) and fully methylated (lane 4) double-stranded oligonucleotides (80 pmol) were incubated with SeqA$_{71-181}$ (200 pmol) at 37°C in the presence of the indicated amounts of salts. After a 30 min incubation, the samples were directly analyzed by 12% poly-acrylamide gel electrophoresis in 1× TBE buffer (90 mM Tris-borate and 2 mM EDTA) run at 3 V/cm for 60 min. Lane 1 indicates a negative control experiment without protein. The DNA bands were visualized by ethidium bromide staining. (**A**) 300 mM NaCl, (**B**) 30 mM NaCl, (**C**) 50 mM NaCl, (**D**) 50 mM KCl, (**E**) 50 mM Li$_2$SO$_4$, (**F**) 50 mM MgSO$_4$ and (**G**) 50 mM MgCl$_2$.
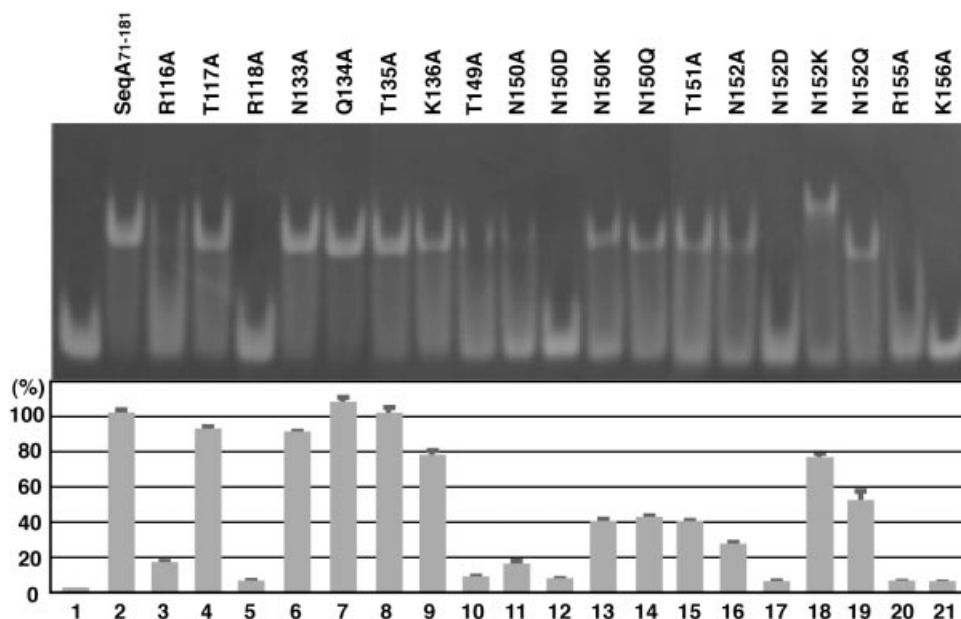
Figure 7. Mutational analysis of SeqA$_{71-181}$. The SeqA$_{71-181}$–DNA interactions were assessed by gel-shift analyses. A hemimethylated DNA substrate containing a G-^{m}A-T-C site at the center was used as the substrate. Lane 1 is a negative control experiment without protein, and lane 2 is a positive control. Graphic representations of the DNA binding by the SeqA$_{71-181}$ mutants are presented in the bottom panel. The amount of complex formation relative to that formed by the wild-type SeqA$_{71-181}$ protein is presented. Average and SD values from three independent experiments are presented. Lanes 3–21 are the experiments with the SeqA$_{71-181}$ mutants, R116A, T117A, R118A, N133A, Q134A, T135A, K136A, T149A, N150A, N150D, N150K, N150Q, T151A, N152A, N152D, N152K, N152Q, R155A and K156A, respectively.

over 19 000 fully methylated G-A-T-C sites in the *E.coli* chromosome, without affecting other methylation-dependent processes, such as mismatch repair and restriction–modification systems.
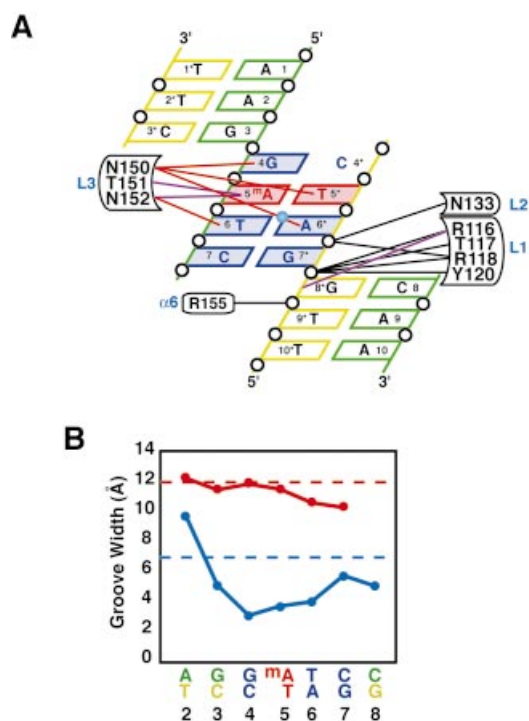
REFERENCES

1. Crooke,E. (1995) Regulation of chromosomal replication in *E.coli*: sequestration and beyond. *Cell*, **82**, 877–880.

Figure 8. The protein–DNA interactions and the DNA distortion in the SeqA$_{71-181}$·DNA complex. (**A**) Schematic diagram summarizing the DNA contacts by SeqA$_{71-181}$. The colors of the DNA correspond to those in Figure 2. Open circles represent phosphates. Hydrogen bonds and salt bridges with the backbone phosphate groups are indicated with black lines. Specific recognitions of bases are shown by red (hydrogen bonds) and purple (van der Waals interactions) lines. (**B**) The widths of the major and minor grooves, calculated by CURVES (31), are plotted against the base number. The red and blue lines indicate the widths of the major and minor grooves, respectively. The red and blue dashed lines indicate the average width of the major and minor grooves, respectively, in the authentic B-form DNA.

2. Henaut,A., Rouxel,T., Gleizes,A., Moszer,I. and Danchin,A. (1996) Uneven distribution of GATC motifs in the *Escherichia coli* chromosome, its plasmids and its phages. *J. Mol. Biol.*, **257**, 574–585.

3. Messer,W., Billekes,U. and Lother,H. (1985) Effect of dam methylation on the activity of the *E.coli* replication origin, *oriC*. *EMBO J.*, **4**, 1327–1332.

4. Messer,W. and Weigel,C. (1996) Initiation of chromosome replication. In Neidhardt,F.C., Curtiss,R.,III, Ingraham,J., Lin,E.C.C., Low,K.B., Magasanik,B., Reznikoff,W.S., Riley,M., Schaechter,M. and Umbarger,H.E. (eds), *Escherichia coli and Salmonella typhimurium: Cellular and Molecular Biology*, 2nd Edn. ASM Press, Washington, DC, pp. 1579–1601.

5. vonFreiesleben,U., Rasmussen,K.V. and Schaechter,M. (1994) SeqA limits DnaA activity in replication from *oriC* in *Escherichia coli*. *Mol. Microbiol.*, **14**, 763–772.

6. Wold,S., Boye,E., Slater,S., Kleckner,N. and Skarstad,K. (1998) Effects of purified SeqA protein on oriC-dependent DNA replication *in vitro*. *EMBO J.*, **17**, 4158–4165.

7. Torheim,N.K. and Skarstad,K. (1999) *Escherichia coli* SeqA protein affects DNA topology and inhibits open complex formation at *oriC*. *EMBO J.*, **18**, 4882–4888.

8. Torheim,N.K., Boye,E., Lobner-Olesen,A., Stokke,T. and Skarstad,K. (2000) The *Escherichia coli* SeqA protein destabilizes mutant DnaA204 protein. *Mol. Microbiol.*, **37**, 629–638.

9. Taghbalout,A., Landoulsi,A., Kern,R., Yamazoe,M., Hiraga,S., Holland,B., Kohiyama,M. and Malki,A. (2000) Competition between the replication initiator DnaA and the sequestration factor SeqA for binding to the hemimethylated chromosomal origin of *E.coli in vitro*. *Genes Cells*, **5**, 873–884.

10. Brendler,T., Abeles,A. and Austin,S. (1995) A protein that binds to the P1 origin core and the *oriC* 13mer region in a methylation-specific fashion is the product of the host *seqA* gene. *EMBO J.*, **14**, 4083–4089.

11. Slater,S., Wold,S., Lu,M., Boye,E., Skarstad,E. and Kleckner,N. (1995) *E.coli* SeqA protein binds *oriC* in two different methyl-modulated reactions appropriate to its roles in DNA replication initiation and origin sequestration. *Cell*, **82**, 927–936.

12. Lu,M., Campbell,J.L., Boye,E. and Kleckner,N. (1994) SeqA: a negative modulator of replication initiation in *E.coli*. *Cell*, **77**, 413–426.

13. Guarne,A., Zhao,Q., Ghirlando,R. and Yang,W. (2002) Insights into negative modulation of *E.coli* replication initiation from the structure of SeqA–hemimethylated DNA complex. *Nature Struct. Biol.*, **9**, 839–843.

14. Fujikawa,N., Kurumizaka,H., Yamazoe,M., Hiraga,S. and Yokoyama,S. (2003) Identification of functional domains of the *Escherichia coli* SeqA protein. *Biochem. Biophys. Res. Commun.*, **300**, 699–705.

15. Brendler,T. and Austin,S. (1999) Binding of SeqA protein to DNA requires interaction between two or more complexes bound to separate hemimethylated GATC sequences. *EMBO J.*, **18**, 2304–2310.

16. Otwinowski,Z. and Minor,W. (1997) Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.*, **276**, 307–326.

17. Hendrickson,W.A. (1991) Determination of macromolecular structures from anomalous diffraction of synchrotron radiation. *Science*, **254**, 51–58.

18. Weeks,C.M. and Miller,R. (1999) The design and implementation of SnB v2.0. *J. Appl. Crystallogr.*, **32**, 120–124.

19. La Fortelle,E. and Bricogne,G. (1997) Maximum-likelihood heavy-atom parameter refinement for multiple isomorphous replacement and multiwavelength anomalous diffraction methods. *Methods Enzymol.*, **276**, 472–494.

20. Collaborative Computational Project Number 4 (1994) The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D*, **50**, 760–763.

21. Jones,T.A., Zou,J-Y., Cowan,S.W. and 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**, 110–119.

22. Brünger,A.T., Adams,P.D., Clore,G.M., DeLano,W.L., Gros,P., Grosse-Kunstleve,R.W., Jiang,J.S., Kuszewski,J., Nilges,M., Pannu,N.S. *et al.* (1998) Crystallography and NMR system: a new software suite for macromolecular structure determination. *Acta Crystallogr. D*, **54**, 905–921.

23. Carson,M. (1991) Ribbons 2.0. *J. Appl. Crystallogr.*, **24**, 958–961.

24. Baikalov,I., Grzeskowiak,K., Yanagi,K., Quintana,J. and Dickerson,R.E. (1993) The crystal structure of the trigonal decamer C-G-A-T-C-^{6me}A-T-C-G: a *B*-DNA helix with 10.6 base-pairs per turn. *J. Mol. Biol.*, **231**, 768–784.

25. Ban,C. and Yang,W. (1998) Structural basis for MutH activation. I. *E.coli* mismatch repair and relationship of MutH to restriction endonucleases. *EMBO J.*, **17**, 1526–1534.

26. Klimasauskas,S., Kumar,S., Roberts,R.J. and Cheng,X. (1994) HhaI methyltransferase flips its target base out of the DNA helix. *Cell*, **76**, 357–369.

27. Reinisch,K.M., Chen,L., Verdine,G.L. and Lipscomb,W.N. (1995) The crystal structure of HaeIII methyltransferase covalently complexed to DNA: an extrahelical cytosine and rearranged base pairing. *Cell*, **82**, 143–153.

28. Goedecke,K., Pignot,M., Goody,R.S., Scheidig,A.J. and Weinhold,E. (2001) Structure of the N6-adenine DNA methyltransferase M*taq*I in complex with DNA and a cofactor analog. *Nature Struct. Biol.*, **8**, 121–125.

29. O'Gara,M., Roberts,R.J. and Cheng,X. (1996) A structural basis for the preferential binding of hemimethylated DNA by HhaI DNA methyltransferase. *J. Mol. Biol.*, **263**, 597–606.

30. Ohki,I., Shimotake,N., Fujita,N., Jee,J.-G., Ikegami,T., Nakao,M. and Shirakawa,M. (2001) Solution structure of the methyl-CpG binding domain of human MBD1 in complex with methylated DNA. *Cell*, **105**, 487–497.

31. Lavery,R. and Skelenar,H. (1988) Definition of generalized helicoidal parameters and an axis of curvature for irregular nucleic acids. *J. Biomol. Struct. Dyn.*, **6**, 63–91.