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

Norie Fujikawa¹, Hitoshi Kurumizaka¹,²,³, Osamu Nureki⁴, Yoshinori Tanaka¹,⁴, Mitsuyoshi Yamazoe⁵, Sota Hiraga⁵ and Shigeyuki Yokoyama¹,²,⁴,*

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

Received as resubmission October 7, 2003; Revised and Accepted November 20, 2003 PDB nos 1IU3 and 1J3E

ABSTRACT

The Escherichia coli SeqA protein recognizes the 11 hemimethylated G-mA-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 N⁶-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-mA-T-C sequences with the minimal DNA-binding domain of SeqA (SeqA71-181), and found that SeqA71-181 specifically binds to hemimethylated DNA containing a sequence with a mismatched mA:G base pair [G-mA(:G)-T-C] as efficiently as the normal hemimethylated G-mA(:T)-T-C sequence. We determined the crystal structures of SeqA71-181 complexed with the mismatched and normal hemimethylated DNAs at 2.5 and 3.0 Å resolutions, respectively, and found that the mismatched mA:G base pair and the normal mA: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 (N⁶-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 N⁶-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-mA-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-mA-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 (SeqA71-181) (14). Consistent with the results from the full-length protein (10,11), SeqA71-181 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 N⁶-methyladenine, is significantly distorted by a base-specific hydrogen bond (13). In contrast, the discrimination mechanism between the
hemimethylated and fully methylated states was not elucidated, 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-mA-T-C sequence are limited to the 5’ three base pairs (G-mA-T) (13), whereas previous biochemical studies showed that the entire G-mA-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 SeqA71-181 to a 14mer DNA containing the hemimethylated G-mA-T-C sequence with a series of mutations. These analyses revealed that all four of the base pairs are characteristically recognized by SeqA71-181. We also unexpectedly found that SeqA71-181 can specifically bind DNA containing a hemimethylated sequence with a mismatched mA:G pair, G-mA(G)-T-C, as strongly as the normal hemimethylated G-mA(T)-T-C sequence. We determined two crystal structures of SeqA71-181 complexed with a 10mer DNA, with either the G-mA(G)-T-C or G-mA(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 N6-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 N6-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 SeqA71-181 protein

The DNA fragment encoding SeqA71-181, which contained amino acid residues 71-181, was ligated into the BamHI site of the pGEX-6P expression vector. SeqA71-181 was expressed in the E.coli strain DH5α as an N-terminal GST fusion protein. The E.coli strain carrying the SeqA71-181 expression vector was grown at 37°C, and isopropyl-β-d-thiogalactopyranoside (IPTG; 1 mM) was added at an OD600 = 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 SeqA71-181 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. SeqA71-181 was recovered in the flow-through fraction, and was further purified on a Superdex 75 HR 10/30 column (Amersham Biosciences). Selenomethionine-labeled SeqA71-181 (SeMet-SeqA71-181) 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 SeqA71-181 (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 electrophoresis 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 SeqA71-181 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 mA:G base pair instead of the mA: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 concentrated up to 5 mg protein/ml, and co-crystals were obtained by the hanging drop method. The crystals of the SeqA71-181-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 P622, 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 SeqA71-181 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 P212121, 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.
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 the 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 flating 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 Rwork of 23.1% and an Rfree 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-mA-T-C sequence in DNA binding by SeqA 71–181

According to the structural analysis by Guarne et al., three base pairs, G:C-mA:T-T:A, within the hemimethylated G-mA-T-C sequence were recognized by the base-specific
interactions with the C-terminal domain of SeqA (13). To test the importance of these base-specific interactions in SeqA binding, we performed a comprehensive mutational study on a 14mer DNA with the hemimethylated G-mA-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 binding (Fig. 1D, lanes 4 and 5). However, the replacement by the G:C base pair significantly reduced the SeqA 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 N6-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 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 SeqA71-181-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 SeqA71-181 complexed with a 10mer hemimethylated T5*G DNA containing the G-mA(G)-T-C sequence (the SeqA71-181-T5*G complex; Fig. 2A) at 2.5 Å resolution (Fig. 2). The 10mer is the minimal size for SeqA71-181 binding of G-mA-T-C DNA (14). The crystals of the SeqA71-181-T5*G complex belong to the orthorhombic space group P212121 (Table 1). We also obtained crystals of SeqA71-181 bound with the normal hemimethylated 10mer DNA containing the G-mA(T)-T-C sequence (the SeqA71-181-T5*G complex; Fig. 2A), 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 SeqA51-181-12mer complex (13)], belong to either the P2221 or P42 space group (Table 1). Nevertheless, the SeqA71-181 structure in the SeqA71-181-T5*G complex is essentially the same as that in the SeqA71-181-10mer complex (r.m.s.d. 0.6 Å; Fig. 2B-D) and the SeqA71-181-12mer complex (r.m.s.d. 0.7 Å).

In the SeqA71-181-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 N6-methyladenine

In the present structure of the SeqA71-181-10mer 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°) [Baikakov et al. (24); Fig. 4A, C and E]. This distortion of the thymine base was also observed in the SeqA51-181-12mer complex (13). In the SeqA71-181-10mer complex, the T5* distortion places the O4 atom of T5* in close enough proximity to form a hydrogen bond with the NH2 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 SeqA71-181-T5*G complex, like the case of T5* in the SeqA71-181-10mer 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-mA-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 this case, 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 SeqAS1-181-12mer 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 5C and 5A). (ii) When the mA5:T5* and T6:A6* base pairs of the SeqAS1-181-10mer complex (Fig. 5B, red) were superimposed on those of the SeqA51-181-12mer complex (Fig. 5B, blue), the mA5: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 SeqAS1-181-12mer complex (Fig. 5B). This difference in the A6* base may be due to the water-mediated hydrogen bonding, which is found in the SeqA71-181-10mer complex but not in the SeqA51-181-12mer complex. (iii) The B-factor of this atom is low (19.78 Å²), 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 SeqA71-181 with different salt concentrations (Fig. 6A and B) and species (Fig. 6C-G). SeqA71-181 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 SeqA71-181 binding to hemimethylated DNA was not affected when the NaCl concentration was reduced to 30 mM (Fig. 6B). In addition, SeqA71-181 specifically bound to the hemimethylated DNA when NaCl was replaced by KCl, Li2SO4, MgSO4 or MgCl2 (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
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 mA: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 SeqA71–181·T5*G and SeqA71–181·10mer 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

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

<table>
<thead>
<tr>
<th>Data collection</th>
<th>SeMet crystal</th>
<th>Native</th>
<th>T5*G mutant</th>
</tr>
</thead>
<tbody>
<tr>
<td>Beamline</td>
<td>BL41XU, SPring-8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wavelength (Å)</td>
<td>0.9798</td>
<td>0.9742</td>
<td>0.9797</td>
</tr>
<tr>
<td>Resolution (Å)</td>
<td>50–3.2</td>
<td>50–3.2</td>
<td>50–3.2</td>
</tr>
<tr>
<td>Unique reflections</td>
<td>13 201</td>
<td>12 864</td>
<td>16 548</td>
</tr>
<tr>
<td>Redundancy</td>
<td>11.9</td>
<td>11.1</td>
<td>11.6</td>
</tr>
<tr>
<td>Completeness (%)</td>
<td>94.6</td>
<td>93.0</td>
<td>98.5</td>
</tr>
<tr>
<td>All data</td>
<td>94.6</td>
<td>93.3</td>
<td>93.4</td>
</tr>
<tr>
<td>Last shell</td>
<td>93.8</td>
<td>89.7</td>
<td>90.9</td>
</tr>
<tr>
<td>All data</td>
<td>8.2</td>
<td>7.3</td>
<td>7.1</td>
</tr>
<tr>
<td>Last shell</td>
<td>2.9</td>
<td>2.8</td>
<td>2.8</td>
</tr>
<tr>
<td>All data</td>
<td>9.3</td>
<td>8.6</td>
<td>9.5</td>
</tr>
<tr>
<td>Last shell</td>
<td>40.2</td>
<td>39.3</td>
<td>40.2</td>
</tr>
<tr>
<td>No. of sites</td>
<td>7</td>
<td>7</td>
<td>7</td>
</tr>
<tr>
<td>Phasing power</td>
<td>0.31</td>
<td>0.53</td>
<td>0.54</td>
</tr>
<tr>
<td>R_{anis} c (%)</td>
<td>0.98</td>
<td>0.94</td>
<td>0.94</td>
</tr>
<tr>
<td>R_{sym} (%)</td>
<td>0.82</td>
<td>0.99</td>
<td>0.89</td>
</tr>
<tr>
<td>R_{free} (%) a</td>
<td>0.30</td>
<td>0.54</td>
<td>0.54</td>
</tr>
<tr>
<td>Resolution (Å)</td>
<td>3.0</td>
<td>2.5</td>
<td></td>
</tr>
<tr>
<td>Protein atoms</td>
<td>1816</td>
<td>1816</td>
<td></td>
</tr>
<tr>
<td>DNA atoms</td>
<td>808</td>
<td>808</td>
<td></td>
</tr>
<tr>
<td>Water molecules</td>
<td>131</td>
<td>95</td>
<td></td>
</tr>
<tr>
<td>R_{work} (%)</td>
<td>23.8</td>
<td>22.6</td>
<td></td>
</tr>
<tr>
<td>Average B-factor (Å²)</td>
<td>79.0</td>
<td>34.2</td>
<td></td>
</tr>
<tr>
<td>Minimum B-factor (Å²)</td>
<td>27.6</td>
<td>7.3</td>
<td></td>
</tr>
<tr>
<td>Maximum B-factor (Å²)</td>
<td>160.6</td>
<td>108.0</td>
<td></td>
</tr>
<tr>
<td>R.m.s.d. bond length (Å)</td>
<td>0.0320</td>
<td>0.00656</td>
<td></td>
</tr>
<tr>
<td>R.m.s.d. bond angles (°)</td>
<td>3.16</td>
<td>1.19</td>
<td></td>
</tr>
<tr>
<td>R.m.s.d. dihedrals (°)</td>
<td>22.8</td>
<td>19.3</td>
<td></td>
</tr>
<tr>
<td>R.m.s.d. impropers (°)</td>
<td>2.67</td>
<td>1.16</td>
<td></td>
</tr>
</tbody>
</table>

aR_{sym} = Σ|I_{avg} - |I_i|/Σ|I_{avg}|
bValues in parentheses represent those for centric reflections.
cR_{anis} = Σ |F_{o}-F_{c}|/Σ|F_{o}|
dR_{free} is calculated for 10% of randomly selected reflections excluded from refinement.
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 δNH₂ 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 δNH₂ 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 SeqA₇₁₋₁₈₁ 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 δNH₂ group of Asn152 (Figs 3C and 8A). When Asn152 was replaced by aspartic acid (N152D), the SeqA₇₁₋₁₈₁ 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 SeqA₇₁₋₁₈₁ 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 18 and 19) of the wild-type binding activity.
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 SeqA71–181 and the 5’-phosphate groups of G8*, G7* and A6*

The SeqA71–181-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 SeqA71–181·10mer 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 SeqA71–181·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 NH2 group of Arg133 interacts with the phosphate group of A6* in the SeqA71–181·10mer 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 SeqA71–181 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-*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-*A-T-C duplex. MutH exhibits structural homology to the PvuII and EcoRV type II endonucleases (25). The mechanism of the hemimethylated G-*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...
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-mA-T-C sequence and the mechanism of its recognition by SeqA provide important insights into how SeqA specifically recognizes the hemimethylated G-mA-T-C sequences among
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.

ACKNOWLEDGEMENTS

We thank Drs M. Kawamoto (JASRI), Y. Kawano (RIKEN)
and N. Kamiya (RIKEN) for help with collecting diffraction
data at SPring-8, and D. Vassylyev (RIKEN, CSL), W.
Kagawa (RIKEN, GSC) and H. Yamaguchi (RIKEN, GSC)
for help with the structure determination and refinement. This
work was supported by the Bioarchitect Research Program
(RIKEN), by the RIKEN Structural Genomics/Proteomics
Initiative (RSGI), and the National Project on Protein
Structural and Functional Analyses, the Ministry of
Education, Sports, Culture, Science, and Technology of Japan.

REFERENCES

Figure 8. The protein–DNA interactions and the DNA distortion in the SeqA71±181-DNA complex. (A) Schematic diagram summarizing the DNA contacts by SeqA71±181. The colors of the DNA correspond to those in the SeqA71±181–DNA complex. (B) Nucleic Acids Research, 2004, Vol. 32, No. 1