

# Structure and Dynamics of Hemimethylated GATC Sites

IMPLICATIONS FOR DNA-SeqA RECOGNITION\*

Received for publication, June 9, 2003

Published, JBC Papers in Press, August 27, 2003, DOI 10.1074/jbc.M306038200

Sung-Hun Bae<sup>‡§</sup>, Hae-Kap Cheong<sup>¶</sup>, Chaejoon Cheong<sup>¶</sup>, Sukhyun Kang<sup>||</sup>, Deog Su Hwang<sup>||</sup>,  
and Byong-Seok Choi<sup>‡\*\*</sup>

From the <sup>‡</sup>Department of Chemistry and National Creative Research Initiative Center, Korea Advanced Institute of Science and Technology, 373-1 Guseong-dong, Yuseong-gu, Daejeon 305-701, Republic of Korea, the <sup>¶</sup>Magnetic Resonance Team, Korea Basic Science Institute, 52 Eoun-dong, Yuseong-gu, Daejeon 305-333, Republic of Korea, and the <sup>||</sup>Department of Microbiology, Seoul National University, Seoul 151-742, Republic of Korea

DNA methylation occurs in most organisms from bacteria to mammals and provides a mechanism for epigenetic control of a variety of cellular processes. In *Escherichia coli*, most of the  $N^6$  positions in adenines found in the sequence GATC are methylated by DNA adenine methyltransferase. After DNA replication, the GATC sites exist transiently in the hemimethylated state, and the specific recognition of these hemimethylated GATC sites is essential for several processes, including sequestration of the site of replication initiation by the SeqA protein, strand discrimination in DNA mismatch repair by the MutH protein, and transcription of several genes. Here, we characterize the solution structure and dynamics of two dodecamer DNA duplexes that each contains a single GATC site in either unmethylated or hemimethylated state. We found that the  $N^6$ -methylated adenine of a hemimethylated GATC site undergoes a slow trans-cis interconversion. The release of a tightly bound cation from hemimethylated DNA explains the instability of this structure. In addition, quantitative structural analysis revealed that hemimethylated DNA has unusual backbone structures and a remarkably narrow major groove. These dynamic and structural features provide insights into the specific recognition of hemimethylated GATC sites by the SeqA protein.

adenines that occur in the sequence GATC are methylated by DNA adenine methyltransferase (2). GATC sites are found more frequently in translated regions of the genome than in non-coding and non-translated sequences, with the exception of the rRNA cluster and some tRNA genes. The spacing between GATC sites never exceeds 2 kb (3). The *E. coli* chromosomal origin of replication (*oriC*) has eleven GATC sites. After DNA replication, the SeqA protein binds specifically to hemimethylated GATC sites within *oriC* and delays remethylation of these sites for 8–10 min (about one third of the cell cycle). This binding prevents multiple initiations of replication within a single round of the cell cycle (4, 5).

In addition to this role in the control of replication initiation, hemimethylated GATC sites constitute a critical factor in the regulation of other cellular processes. GATC sites that are located within chromosomal sequences other than *oriC* have a much shorter hemimethylation half-life (2–4 s) after replication (6). In the process of DNA mismatch repair, the transient presence of hemimethylated GATC sites allows the repair protein MutH to specifically recognize and cleave only the newly synthesized unmethylated DNA strand (7). The transcription of genes such as *dnaA2P*, *mioC*, *trpS*, and *trpR* is induced when their promoters are transiently hemimethylated after replication (8, 9). The same strategy is utilized in the transcription of certain transposons (10).

The crystal structure of a single  $N^6$ -methylated adenine base showed that the methyl group attached at the adenine  $N^6$  position points toward the hydrogen bonding interface of the Watson-Crick base pair (the cis form, Fig. 1A) (11). This study led to the suggestion that methylation of adenine  $N^6$  would disrupt the secondary structure of the DNA double helix (11, 12). However, structural studies of double-stranded oligonucleotides revealed that  $N^6$ -methylated adenine forms a normal Watson-Crick base pair with the thymine in the complementary DNA strand (with the trans form, Fig. 1A), and there is no significant conformational difference between methylated DNA and unmethylated DNA with the same nucleotide sequence. These studies, however, demonstrated that the adenine  $N^6$  hemimethylated DNA has a lower melting temperature and distinct dynamic characteristics represented by slow duplex-single strand exchange (13–16). Here, we investigate the solution structure and dynamics of two dodecamer DNA duplexes that contain either a single hemimethylated or unmethylated GATC site. Our study revealed that the  $N^6$ -methylated adenine of hemimethylated GATC undergoes a slow interconversion between the trans and cis forms without significant distortions of the double helix and that the instability of hemimethylated DNA is correlated with the easy release of a tightly bound cation at an elevated temperature. In addition to this unusual

DNA methylation occurs in organisms as diverse as bacteria and mammals and provides a mechanism for the epigenetic control of the timing and targeting of several types of cellular functions (1). The sites commonly methylated by DNA methyltransferases are the  $C^5$  and  $N^4$  positions of cytosine and the  $N^6$  position of adenine. Most DNA methyltransferases recognize palindromic DNA sequences, and methylated DNA sites can exist in the unmethylated, hemimethylated, or fully methylated states (for example, GATC/GATC, GATC/GA<sup>m</sup>TC, or GA<sup>m</sup>TC/GA<sup>m</sup>TC, respectively, with A<sup>m</sup> representing  $N^6$ -methylated adenine). In *Escherichia coli*, most of the  $N^6$  positions of

\* This work was supported by the Creative Research Initiative from the Ministry of Science and Technology, the Republic of Korea. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

The atomic coordinates and structure factors (codes 1OPQ, 1OQ2, and 1UAB) have been deposited in the Protein Data Bank, Research Laboratory for Structural Bioinformatics, Rutgers University, New Brunswick, NJ (<http://www.rcsb.org/>).

§ Partially supported by the BK21 project.

\*\* To whom correspondence should be addressed. Tel.: 82-42-869-2828; Fax: 82-42-869-2810; E-mail: byongseok.choi@kaist.ac.kr.

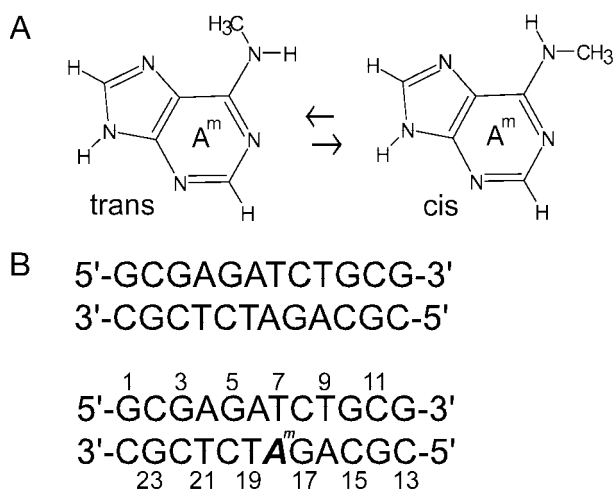


FIG. 1. A, the trans and cis forms of  $N^6$ -methylated adenine. B, sequence and numbering of the dodecamer DNA duplexes used in this study (numbering of the unmethylated DNA is identical).  $A^m$  indicates the  $N^6$ -methylated adenine.

dynamics, quantitative structural analysis revealed that a hemimethylated GATC site has unusual metastable backbone structures and that the major groove width of a hemimethylated GATC site is remarkably narrow. These dynamic and structural features provide insights into the specific binding of hemimethylated GATC sites by SeqA.

#### EXPERIMENTAL PROCEDURES

**Sample Preparation**—All DNA oligonucleotides were synthesized on a solid support with an Applied Biosystems 391 DNA synthesizer using standard phosphoramidite chemistry. Adenine  $N^6$ -methylated strands were synthesized by using  $N^6$ -methylated 5'-dimethoxytrityl- $N$ -benzoyl-2'-deoxyadenosine,3'-((2-cyanoethyl)-(N,N-diisopropyl))-phosphoramidite (dA-CE phosphoramidite; Glen Research, Inc.). The buffer conditions for NMR experiments were 10 mM sodium phosphate (pH 6.8) and 100 mM sodium chloride unless otherwise specified.

**NMR Experiments**—NMR spectra were acquired on a Varian Inova 600 MHz spectrometer. Two-dimensional nuclear Overhauser effect (NOE)<sup>1</sup> spectroscopy (NOESY) ( $\tau_m = 240$  ms) was carried out in 95%  $H_2O$ , 5%  $D_2O$  at 4 and 12 °C. Two-dimensional NOESY ( $\tau_m = 60, 120,$  and 240 ms), two-dimensional double quantum filter correlation spectroscopy (DQF-COSY), two-dimensional total correlation spectroscopy (TOCSY) ( $\tau_m = 80$  ms),  $^1H$ - $^{31}P$  heteronuclear correlation spectroscopy, and rotational NOE spectroscopy (ROESY) ( $\tau_m = 240$  ms) were conducted in 99.96%  $D_2O$  at 30 °C. One-dimensional natural abundance  $^1H$ - $^{13}C$  HSQC spectra were obtained at 0, 6, 12, 30, 42, 48, 54, 60, and 66 °C with selective carbon pulses for the adenine  $N^6$ -methyl group.

**Base Pair Exchange Kinetics**—Selective longitudinal relaxation times were measured at increasing concentrations of  $NH_3$  ranging from 0 to 0.6 M at pH 8.8 and 15 °C. Selective inversion was achieved by an IBURP2-shaped (17) 180 pulse with a 1.5-kHz bandwidth centered at 12.96 ppm. The water signal was suppressed by a jump-and-return pulse sequence. Interpretation and analysis of the data followed the previously reported methods (18).

**DNA Binding of  $NH_4^+$  Ions**—DNA binding of  $NH_4^+$  ions was probed by one-dimensional  $^{15}N$ -decoupled  $^1H$ - $^{15}N$  HSQC spectra, which were acquired with ~2 mM DNA, 10 mM sodium phosphate (pH 6.8), and 20 mM  $^{15}NH_4Cl$ . The  $^{15}N$  carrier was centered at 17.5 ppm, and the  $^1H$  sweep width was 20 ppm. 32 scans of 32768 complex points were acquired for each spectrum.

**Structure Calculation**—The distance constraints were derived from the integrated NOE volumes and three assumed isotropic correlation times ( $\tau_c = 3, 4,$  and 5 ns) using a relaxation matrix analysis program, MARDIGRAS (19). The  $\delta$  dihedral angle was derived from the analysis of  $^3J_{H1'-H2'}$  in double quantum filter correlation spectroscopy.  $\chi$  was constrained to  $220 \pm 45^\circ$  based on the medium-to-weak intra-residue H6/H8-H1' NOE.  $\alpha$  and  $\zeta$  were unconstrained, and other backbone

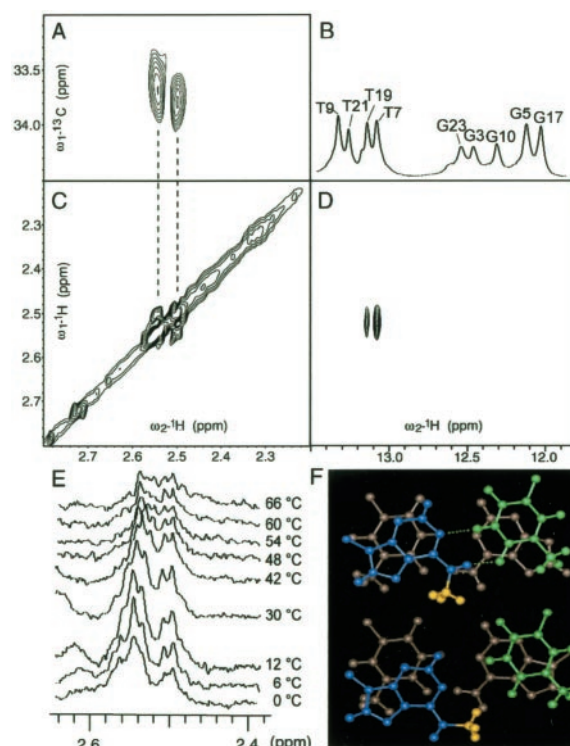


FIG. 2. **Trans-cis interconversion of the  $N^6$ -methylated adenine.** A, section of two-dimensional  $^1H$ - $^{13}C$  HSQC spectrum. B, one-dimensional spectrum of imino protons. The chemical shift is aligned to  $\omega_2$  of the spectrum shown in panel D. C, section of two-dimensional rotational NOE spectroscopy spectrum ( $\tau_m = 240$  ms). D, section of two-dimensional  $H_2O$  NOESY spectrum ( $\tau_m = 240$  ms). The chemical shifts of  $\omega_2$  of panel A and  $\omega_1$  of panel D are aligned to that of  $\omega_2$  and  $\omega_1$  of the spectrum of panel C, respectively. E, one-dimensional  $^1H$ - $^{13}C$  HSQC spectra with increasing temperature from 0 to 66 °C. Chemical shift is shown for the spectrum acquired at 30 °C. Both the trans and cis peaks (not specifically assigned) are split by a scalar coupling between  $N^6$ - $CH_3$  and  $N^6$ -H (~6 Hz). F, stacked view of the central base pairs, T7 (green)-A18 (blue)/A6 (gray)-T19 (gray), of the average structures of the trans (upper) and the cis form (lower) hemimethylated DNA. Dotted lines indicate hydrogen bonding.  $N^6$ -methyl group is shown in yellow.

dihedral angles were loosely constrained to the standard B-form ( $\beta$  ( $180 \pm 30^\circ$ ),  $\gamma$  ( $60 \pm 30^\circ$ ), and  $\epsilon$  ( $230 \pm 70^\circ$ )), except that  $\gamma$  was unconstrained for the hemimethylated DNA. Scalar  $^1J_{CH}$  and dipolar  $^1D_{CH}$  couplings were derived from natural abundance  $^1H$ - $^{13}C$  HSQC experiments with and without Pfl (~10 mg/ml) at 30 °C. Pfl filamentous bacteriophage was purchased from ASLA, Ltd. Alignment tensor analysis of the observed residual dipolar couplings was performed as described previously (20). All structure calculations were performed by crystallography and NMR system (CNS) software (21). Two extended strands were used as the starting structure, which was subjected to 60 ps of torsion angle dynamics at 20,000 K, followed by 60 ps of torsion angle dynamics cooling to 0 K and 40 ps of Cartesian dynamics cooling from 3,000 to 0 K. The final structures were generated after 2,000 cycles of energy minimization. The distance force constant was 50 kcal  $mol^{-1} \cdot \text{\AA}^{-2}$ , and the dihedral angle force constant, which initially was 5, was scaled to 250 kcal  $mol^{-1} \cdot \text{radian}^{-2}$  during cooling. The force constant for the residual dipolar coupling energy function was 3 kcal  $mol^{-1}$ . For unmethylated DNA, 18 of 100 trial structures were converged. For hemimethylated DNA, each set of structure calculations was performed for the trans and cis isomers (selected converged structures were 16 of 100 for both). The final structures were analyzed by MOLMOL (22), 3DNA (23), and Curves 5.2 (24) software.

#### RESULTS

**Identification of  $^1H$  and  $^{13}C$  Resonances of  $N^6$ -Methyl Group**—In the  $H_2O$ -NOESY ( $\tau_m = 240$  ms) of hemimethylated DNA, T19 and T7 imino resonances showed medium intensity NOEs with a ~2.5 ppm resonance (Fig. 2B and D). This chemical shift is usually associated with H2' and H2'' resonances; however, the distance between the imino proton and the H2'/

<sup>1</sup> The abbreviations used are: NOE, nuclear Overhauser effect; NOESY, NOE spectroscopy; HSQC, heteronuclear single quantum correlation spectroscopy.

H2" protons is  $> \sim 5$  Å in normal A- or B-form DNA helices. Consistently, no NOE was observed in the corresponding region of the H<sub>2</sub>O-NOESY obtained with the unmethylated DNA (data not shown). Natural abundance <sup>1</sup>H-<sup>13</sup>C HSQC showed that this proton is attached to a carbon whose <sup>13</sup>C chemical shift is  $\sim 34$  ppm, which is an intermediate between sugar C2' ( $\sim 40$  ppm) and T-methyl carbon ( $\sim 15$  ppm) chemical shifts. Improved resolution of D<sub>2</sub>O <sup>1</sup>H-<sup>13</sup>C HSQC spectra showed that this cross-peak actually consists of two cross-peaks (Fig. 2A) and that the cross-peak between them has the same sign as the diagonal peak in a D<sub>2</sub>O rotational NOE spectroscopy spectrum ( $\tau_m = 240$  ms) (Fig. 2, A and C), implying that this cross-peak originated from the exchange of conformations with two different magnetic environments. Therefore, we reasonably assigned these two distinct resonances as the methyl resonances of the trans and cis forms of N<sup>6</sup>-methylated adenine. In addition, previous NMR studies performed at the mononucleotide level support this resonance assignment. For example, an NMR study of N<sup>6</sup>-methylated adenine in CDCl<sub>3</sub> solvent showed that the N<sup>6</sup>-methyl chemical shifts are 3.07 (cis) and 3.51 (trans) ppm (25). Another study of an N<sup>6</sup>-methylated adenine monomer in H<sub>2</sub>O observed that the methyl resonance is a doublet and that the amino proton resonance is a spin-coupled quartet (26).

**Slow Trans-Cis Interconversion of the N<sup>6</sup>-Methylated Adenine**—A series of N<sup>6</sup>-methyl-selective natural abundance <sup>1</sup>H-<sup>13</sup>C HSQC experiments performed at increasing temperatures (0–66 °C) revealed that the trans-cis interconversion of the N<sup>6</sup>-methylated adenine via C<sup>6</sup>-N<sup>6</sup> bond rotation is quite slow (Fig. 2E). From 0 to 12 °C the separation of the two peaks remained at  $\sim 24$  Hz, and the separation decreased to  $\sim 15$  Hz for increasing temperatures up to 48 °C. However, even at temperatures above 48 °C the coalescence was not observed. If we assume that the initial separation between these two resonances in the absence of C<sup>6</sup>-N<sup>6</sup> bond rotation is 24 Hz, the simulated exchange kinetics estimates that the exchange process should occur  $< 10$  times per second (s<sup>-1</sup>) throughout the temperature range that we investigated. This slow exchange is consistent with the NOESY spectra in which many resonances have line widths that are broader than the corresponding resonances of the unmethylated DNA, and a few resonances have two sets of chemical shifts (data not shown).

The C-N bond rotation of an adenine amino group in a double-stranded helix has been shown to be 60–24,000 s<sup>-1</sup> over temperatures ranging from 0 to 70 °C (27). Because the C<sup>6</sup>-N<sup>6</sup> bond length of N<sup>6</sup>-methylated adenine (1.342 Å) (11) is nearly identical to that of a normal adenine C<sup>6</sup>-N<sup>6</sup> bond length (1.335 Å), the slow exchange of the N<sup>6</sup>-methylated adenine should be explained by structural reasons other than the partial double bond character of the C<sup>6</sup>-N<sup>6</sup>.

**Overall Structures of Unmethylated and Hemimethylated DNAs**—For both unmethylated and hemimethylated DNA, NOE connectivity and sugar puckers (C2'-endo, <sup>3</sup>J<sub>H1'-H2'</sub>  $> 8$  Hz) typical of B-DNA were observed, and the <sup>31</sup>P chemical shifts ( $\sim -3$  ppm) were populated within a small range. However, one deviation from the standard B-DNA structure was that the minor groove width of unmethylated DNA was significantly narrow (Fig. 3A). This structural feature of the GATC site is consistent with the crystal structure of the self-complementary dodecamer DNA duplex, 5'-CGTAGATCTACG-3' (28), where the spine of hydration is observed in the narrow minor groove.

**Trans and Cis Structures of the Hemimethylated GATC Site**—Even though the N<sup>6</sup>-methylated adenine undergoes a slow interconversion between two distinct isomers, NOESY spectra showed only a single set of cross-peaks, except for a few

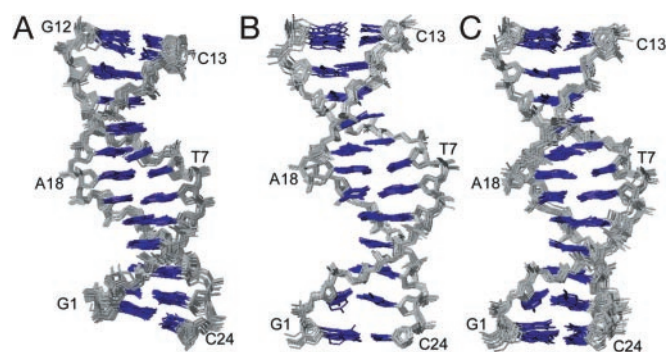


FIG. 3. Superimposed overall structures of unmethylated DNA (panel A; 18 structures) and the trans form (panel B; 16 structures) and cis form (panel C; 16 structures) of hemimethylated DNA.

resonances. Therefore, to explore the small differences between the two isomeric GATC sites, two separate calculations were performed for the trans and cis forms of hemimethylated DNA. For these calculations, most of the same constraints were used, except that hydrogen bonding restraints between A18 and T7 were used for the trans version but not for the cis one. Unlike for the unmethylated DNA, the backbone dihedral angle  $\gamma$  was unconstrained for the hemimethylated DNA (in both the trans and cis structures), because the normal  $\gamma$  ( $60 \pm 30^\circ$ ) angle constraints yielded results that consistently violated other NOE constraints.

The structural changes between the trans and cis isomers were not extensive, and backbone dihedrals showed no remarkable differences between the trans and cis isomers. Also, the sugar puckers of hemimethylated DNA were pure C2'-endo and not intermediate between C2'-endo and C3'-endo, implying that the sugar puckers were not altered by the interconversion. The root mean square deviation between the trans and cis isomers was  $1.1 \pm 0.2$  Å, which is slightly larger than the pairwise root mean square deviation values of converged sets of the trans ( $0.6 \pm 0.2$  Å) and cis ( $0.9 \pm 0.2$  Å) isomers (Table I). The stacking interactions of the GATC site are quite similar for the trans and cis isomers, except that the stacking between A18 and T19 is much stronger in the trans isomer (Fig. 2F). Despite the trans-cis isomerization of the N<sup>6</sup>-methylated adenine (A18), the observed NOE between A18H2 and T7 imino was of strong intensity, which is typical of the Watson-Crick A-T base pairs. This NOE constraint was used identically for the trans and cis structures and did not violate other NOE and angle constraints, even in the cis structures. The distance between the A18H2 and the T7 imino in the cis structures was slightly shorter ( $2.88 \pm 0.02$  Å) than that in the trans structures ( $3.14 \pm 0.04$  Å) (Fig. 2F). This finding suggests that steric clash between the cis form of the N<sup>6</sup>-methylated adenine (A18) and the complementary strand's base (T7) is relieved by the shift of the A18 base into the major groove without the complete opening of the base pair.

**Characteristic Structural Identities of the Hemimethylated GATC Site in Regard to the Backbone**—For hemimethylated DNA, the  $\alpha/\gamma$  angles were correlated as *gauche*+/*gauche*- in residues A4, A6, A16, A18, and C20. The  $\gamma$  angle of C8 was *trans* for trans hemimethylated DNA and spanned the *gauche*- to *trans* range for cis hemimethylated DNA. Molecular simulations of a GpC step of a number of DNA crystal structures showed that, in addition to the canonical  $\alpha/\gamma$  angles (*gauche*-/*gauche*+), a few metastable sub-states, including  $\alpha/\gamma$ :*gauche*+/*gauche*-, exist in DNA-protein complexes and free DNA (29). This  $\alpha/\gamma$ :*gauche*+/*gauche*- sub-state is accompanied by a BI conformation, which is characterized by an  $\epsilon/\zeta$ :*trans*/*gauche*- conformation with the ( $\epsilon$ - $\zeta$ ) around  $-90^\circ$ . The

TABLE I  
Structure determination statistics for  
unmethylated and hemimethylated DNA

	Unmethylated	Hemimethylated	
		Cis	Trans
Total number of NOE distance restraints	438	409	
Intra-residue distances	103	97	
Sequential residue distances	232	218	
Inter-strand distances	103	94	
Dihedral restraints ( $\beta$ , $\gamma$ , $\delta$ , $\epsilon$ , and $\chi$ )	113	90	
Base planarity restraints	12	12	
Residual dipolar coupling restraints	20	27	
Total number of restraints	583	538	
Pairwise r.m.s.d for all heavy atoms (Å) <sup>a</sup>	0.8 ± 0.2	0.9 ± 0.2	0.6 ± 0.2
Average NOE violations (Å)	0 (>0.5 Å)	0 (>0.5 Å)	0 (>0.5 Å)
Average dihedral angle violations (°)	0 (>5°)	0 (>5°)	0 (>5°)
Mean deviation from covalent geometry			
Bond lengths (Å)	0.0018	0.0018	0.0018
Angles (°)	0.4	0.4	0.4
Improper (°)	0.4	0.6	0.5

<sup>a</sup> r.m.s.d. is root mean square deviation.

DNA backbones of both the unmethylated and hemimethylated DNA were BI conformations throughout all the residues.

**Characteristic Structural Identities of the Hemimethylated GATC Site in Regard to Groove Width**—The minor groove width measured from P-P distances gets progressively narrower from the 5' and 3' ends of the oligomer to the central GATC site of unmethylated DNA (Figs. 3A and 4). However, hemimethylated DNA shows a quite different pattern with respect to the minor groove width. Residues 5' to the N<sup>6</sup>-methylated adenine (A18) have narrower minor groove widths, whereas residues on the 3' side have wider minor groove widths than those observed for unmethylated DNA, and this is true for both the cis and trans structures (Figs. 3, B and C, and 4). The distances between the adenine H2 ( $n$ ) and the sequential ( $n + 1$ ) or inter-strand H1' ( $n' + 1$ ) ( $n$  and  $n'$  are complementary residues) are dependent on the local sequence and show a very good correlation with minor groove parameters P-P or H1'-H1' of DNA (30). For unmethylated DNA, the NOEs of A18H2-T19H1', A6H2-T7H1', A18H2-C8H1', and A6H2-C20H1' are of medium-to-weak intensity. However, for hemimethylated DNA, these NOEs are of weak-to-very weak intensity, which is quite consistent with the obtained structures, given that the widened minor groove on the 3' side of A18 makes these distances longer by ~1–2 Å.

The major groove width of unmethylated DNA gets wider as one progresses inward from the 5' and 3' ends to a maximum of  $13.6 \pm 0.7$  Å at the central A6-T7/A18-T19 base pair step (Fig. 4). However, the major groove width of hemimethylated DNA shows sharply contrasting results. With a minimum ( $10.1 \pm 0.9$  Å for the cis and  $10.9 \pm 0.2$  Å for the trans) at the A6-T7/A18-T19 base pair step, the major groove width widens as one progresses outward toward the 5' and 3' ends (Fig. 4).

This unusual helical feature could be explained partially by the twist angles. Unmethylated DNA shows typical twist angles ( $35 \pm 1^\circ$ ) through the G5-A6/T19-C20 base pair step to the C8-T9/A16-G17 base pair step. However, the T7-C8/G17-A18 base step of hemimethylated DNA is over-twisted ( $46 \pm 2^\circ$  for the cis and  $40 \pm 1^\circ$  for the trans), and the A6-T7/A18-T19 base step is under-twisted ( $23 \pm 2^\circ$  for the cis and  $30 \pm 1^\circ$  for the

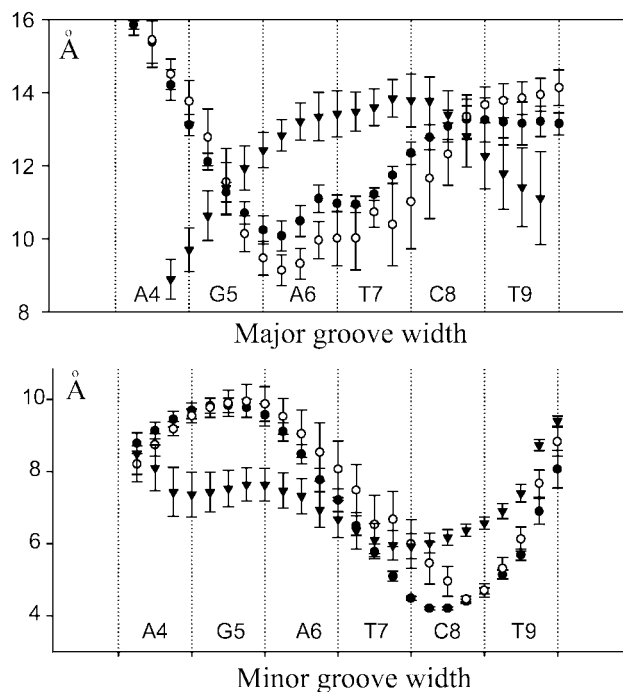


FIG. 4. Major and minor groove widths of the converged structures of unmethylated (triangle) and hemimethylated DNAs (cis, open circle; trans, filled circle). Groove widths were calculated by Curves 5.2 (24). Error bars indicate S.D.

trans). The wider major groove width expected for the 5' side, and the narrower major groove width expected for the 3' side of the A18 residue of hemimethylated DNA might be adapted to the current groove widths by the over-twisted and under-twisted helical distortions.

**Base Pair Lifetimes of the Hemimethylated GATC Site**—We measured the base pair lifetimes of the central A-T base pairs in unmethylated and hemimethylated GATC sites using ammonia as a base catalyst. In the helical state, the T imino protons are protected from exchange with base catalyst, but under a higher concentration of base catalyst an exchange may take place each time a base pair opens. Because the reported lifetimes of typical A-T and G-C base pairs are 0.5–7 and 4–50 ms, respectively (18), the lifetime we measured for the A6-T19 and A18-T7 base pairs (the T7 and T19 imino resonances became broad and overlapped as the ammonia concentration increased) in an unmethylated GATC site (5.8 ms) is a reasonable value (Fig. 5A). However, the lifetime we measured for these base pairs in a hemimethylated GATC site (20.4 ms) is remarkably longer than expected (Fig. 5A). These findings support previous findings of base pair lifetimes in GATC sites, even though the previously adopted methods differ from ours (15).

The longer base pair lifetime of hemimethylated DNA might be caused by the fact that the ammonium ion can bind to the minor groove of DNA and affect base pair kinetics without altering the DNA structure (31). Thus, we assessed ammonium ion binding with the use of <sup>1</sup>H-<sup>15</sup>N HSQC of a <sup>15</sup>N-labeled ammonium ion (NH<sub>4</sub><sup>+</sup>) in which the <sup>1</sup>H-<sup>15</sup>N correlation peak will be observed only for the ammonium ions that are tightly bound to DNA (32). This is because free bulk ammonium ions in solution experience extensive proton exchange with solvent water molecules. In contrast to the control experiment without DNA, wherein no <sup>1</sup>H-<sup>15</sup>N correlation peak was observed, both unmethylated and hemimethylated DNA showed a clear <sup>1</sup>H-<sup>15</sup>N correlation peak at least at the temperature at which we measured the base pair lifetimes (12 °C) (Fig. 5C). Therefore,

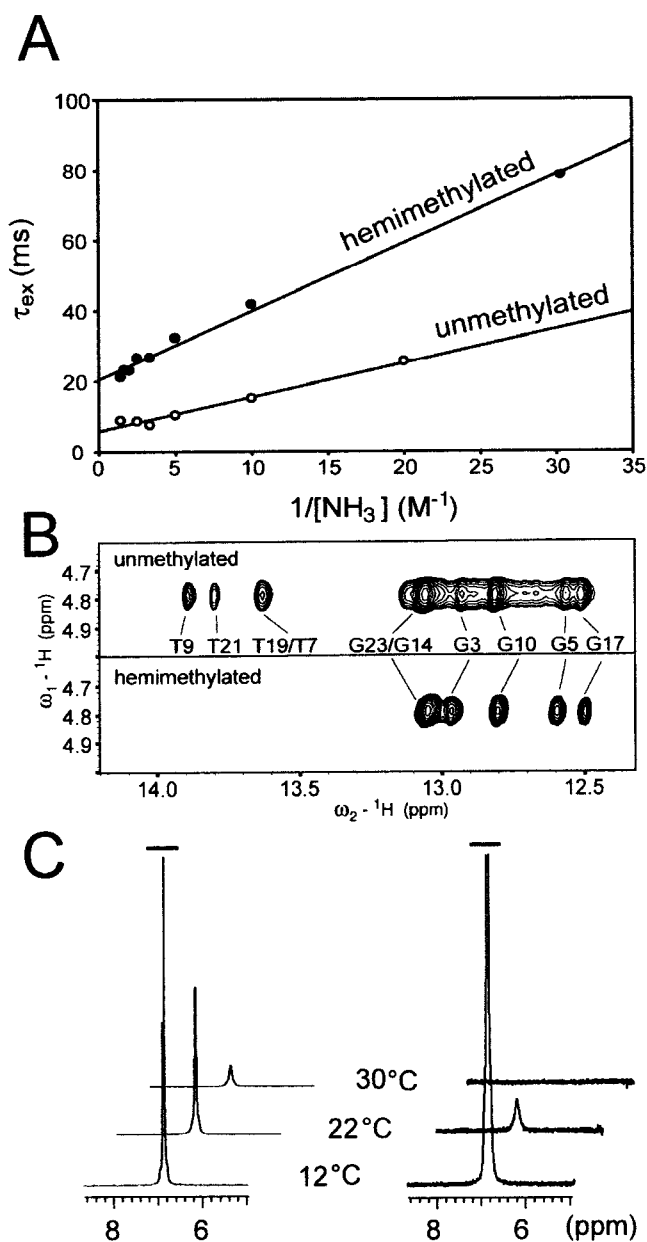


FIG. 5. Dynamics of hemimethylated GATC site. A, exchange time of the imino proton of T7 (and T19) at pH 8.8, 15 °C. Filled circle, hemimethylated DNA; open circle, unmethylated DNA. B, section of H<sub>2</sub>O NOESY spectrum ( $\tau_m = 240$  ms). The chemical shift of water resonance is  $\sim 4.8$  ppm. C, one-dimensional  $^1\text{H}$ - $^{15}\text{N}$  HSQC spectra of  $^{15}\text{NH}_4\text{Cl}$  in the presence of unmethylated and hemimethylated DNA. The first spectra (12 °C) of both DNAs are scaled to same peak height. Only half of the peak is shown for the first spectrum because of space limitation.

the effect of ammonium ions appears to be similar for unmethylated and hemimethylated DNA, so that the longer base pair lifetime observed for hemimethylated GATC site is not the result of differential  $\text{NH}_4^+$  binding.

There are a number of possible reasons for the observed difference in base pair lifetimes in unmethylated and hemimethylated GATC sites. First, the opening of a base pair facilitates the trans-to-cis isomerization of the  $N^6$ -methylated adenine, which should be accompanied by steric clashes at the major groove. In some cases, the bases would return to the closed position before the base pair is able to open completely, and this will slow down the exchange of the imino proton. Second, the major groove width of the central part of the hemimethylated DNA is 1–2 Å narrower than that of the un-

methylated DNA, which could be unfavorable for the exchange of the imino proton or for the approach of a base catalyst to the major groove. The reduced exposure of T imino protons is supported by the H<sub>2</sub>O NOESY spectra, where the cross peaks from the NOE or exchange between water protons and T imino protons are absent for the hemimethylated DNA (Fig. 5B).

**Instability and Cation Binding of Hemimethylated DNA**—The melting temperature of hemimethylated DNA containing a GATC site is lower than unmethylated DNA with same sequence by  $>5$  °C (13, 33). This fact appears to be inconsistent with the longer base pair lifetime that we observed for the hemimethylated GATC site. However, the  $^1\text{H}$ - $^{15}\text{N}$  HSQC experiments with the  $^{15}\text{N}$ -labeled ammonium ion at increasing temperatures yielded interesting clues to the differential stability of unmethylated and hemimethylated GATC sites (Fig. 5C). For both unmethylated and hemimethylated DNA, a clear  $^1\text{H}$ - $^{15}\text{N}$  correlation was observed at 12 °C, but the signal decreased as the temperature increased, implying that more ammonium ions are getting exposed to bulk solvent water. At 30 °C, no signal remained for hemimethylated DNA, whereas a residual signal ( $\sim 10\%$  of 12 °C) existed for unmethylated DNA. The hemimethylated DNA released tightly bound cation more easily than did the unmethylated DNA, and, therefore, the hemimethylated DNA was more easily destabilized because binding of a mono- or a di-valent cation is important for the stabilization of negatively charged DNA molecules.

#### DISCUSSION

It is clear that the absence or presence of a methyl group at the  $N^6$  position of adenine in GATC sites affects the specificity of DNA-protein recognition. Proteins that are designed to detect hemimethylated GATC sites probably function in at least two cooperative ways. First, these proteins might directly recognize that a methyl group has been substituted for the hydrogen at the  $N^6$  position of the adenine residue. Second, these same proteins might recognize methylation-induced structural and/or subsequent dynamic changes in the GATC region of the double helix.

A number of proteins are known to interact exclusively with hemimethylated GATC sites. The SeqA protein, for which the protein-DNA complex structure is available, recognizes hemimethylated GATC sites in the background of fully methylated GATC sites. However, *in vitro* data indicate that SeqA also discriminates between hemimethylated and unmethylated GATC sites (34, 35). From the crystal structure of the SeqA-hemimethylated DNA complex, we know that SeqA binds to the major groove of hemimethylated DNA and contacts the adenine  $N^6$ -methyl group of the GATC site via van der Waals interactions of the protein backbone between Thr-151 and Asn-152. However, the sequence-specific interactions are limited to the central A-T base pairs (34). *In vitro* studies have shown that substitution of either of the G-C base pairs in the GATC site inhibits binding of SeqA (36), implying that additional factors participate in the determination of binding specificity.

The structure of the SeqA-hemimethylated DNA complex revealed that the phosphate backbone of the unmethylated strand fits into the positively charged groove formed by Arg-116 and Arg-155, and that of the methylated strand fits into the more loose groove enclosed by Arg-86 (34) (Fig. 6A). In our SeqA-DNA model complexes (Fig. 6A), the trans-hemimethylated DNA fits reasonably well into the grooves of SeqA, and the  $N^6$ -methyl group is positioned at a location similar to that in the original SeqA-hemimethylated DNA complex. However, the backbones of the unmethylated DNA clash with the two groove structures of SeqA, because the major groove of this DNA structure is larger than the DNA structure of the original SeqA-hemimethylated DNA complex by  $>1$  Å (Fig. 6B). This

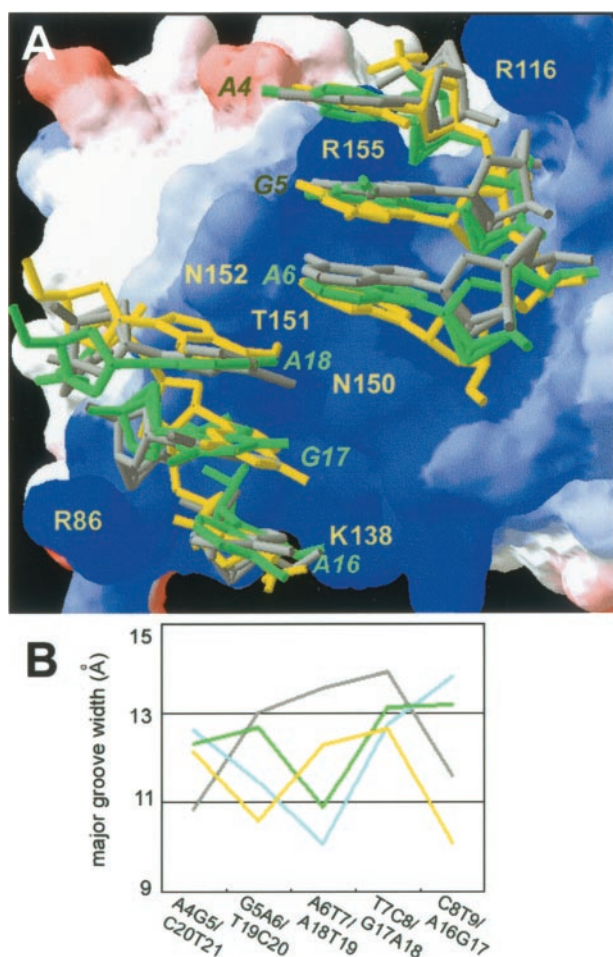


FIG. 6. Recognition of hemimethylated GATC site. A, modeled structures of SeqA-DNA complexes. These models were constructed with MOLMOL (22) by fitting the backbones of the DNA residues 4–6, 16, and 17 of our structures (unmethylated and trans-hemimethylated DNA) with those of the crystal structure of the SeqA-hemimethylated DNA complex (34) (PDB code 1LRR; polymer chains A–C). Molecular surface of SeqA protein was generated with a Swiss-pdb viewer (39). Positive and negative protein residues are shown in blue and red, respectively. Selected DNA residues of the original crystal structure are shown in yellow, and those of hemimethylated and unmethylated DNA structures are shown in green and gray, respectively. R86, Arg-86; R116, Arg-116; K138, Lys-138; N150, Asn-150; T151, Thr-151; N152, Asn-152; and R155, Arg-155. B, major groove width of hemimethylated DNA (trans is green and cis is cyan), unmethylated DNA (gray), and the DNA of SeqA complex (yellow) (34).

suggests that the major groove width reduced by adenine  $N^6$ -methylation should be an important factor for the binding specificity.

The  $N^6$ -methylated adenine within a DNA double helix alternates between the two distinct conformations on an NMR time scale. However, the  $N^6$ -methyl group has been shown to point away from the hydrogen bonding interface of the Watson-Crick base pair (the trans form) in the crystal (37) and solution (13, 16) structures of hemimethylated (13, 37) or fully methylated (16) DNA with varied sequences. The discrepancy between the previous results and our current findings can be understood when we consider that these previous studies lacked dynamic information on the trans-cis interconversion. In the structure calculations using NMR spectroscopy, the NOE that we and previous studies observed between H2 of the  $N^6$ -methylated adenine and the imino proton of complementary strand's thymine will constrain this base pair as the trans form unless the potential existence of the cis form is considered. In actuality, this NOE is consistent with both the trans and cis

forms. With regard to the crystal structure, the  $N^6$ -methylated adenine might be induced to have the trans form during crystallization, because the trans form has more favorable base stacking and hydrogen bonding interactions.

The trans-cis interconversion might be accompanied by a transient helical kink toward the minor groove, which can assist the rotation of the bulky  $N^6$ -methyl group in the major groove. We suggest that the trans-cis interconversion and inherent instability of the hemimethylated GATC site could facilitate the specific recognition of this site because of the potential for induced fitting of DNA structure. In the complex structure of SeqA and hemimethylated DNA, DNA is slightly kinked ( $8^\circ$ ) toward the minor groove (34). Also, the footprinting experiment of hemimethylated DNA, which is bound by SeqA, has shown that the footprint of the methylated strand is broader than that of the unmethylated strand (36). Similarly as with the case of DNA footprinting analysis of TATA binding protein (38), this asymmetry could be understood as a deformation of DNA structure induced by SeqA binding.

In conclusion, our data showed that the  $N^6$ -methylated adenine of a hemimethylated GATC site slowly interconverts between the trans and cis forms. This interconversion did not alter the DNA double helical structure significantly. We confirmed that the base pair lifetimes of the central A-T base pairs in the hemimethylated GATC site are unusually long and that this result is not caused by differential binding of the base catalyst to the hemimethylated DNA. In addition, we showed that the hemimethylated GATC site loses its tightly bound cation more easily than does the unmethylated GATC site, which explains the instability of the hemimethylated GATC site. Together with the dynamic features, the structural results of the metastable backbone and the unique groove width patterns, including narrow major groove width, provide a basis for understanding the specific recognition of hemimethylated GATC sites by SeqA protein.

#### REFERENCES

- Jeltsch, A. (2002) *ChemBiochem* **3**, 274–293
- Lacks, S., and Greenberg, B. (1977) *J. Mol. Biol.* **114**, 153–168
- Barras, F., and Marinus, M. G. (1988) *Nucleic Acids Res.* **16**, 9821–9838
- Ogden, G. B., Pratt, M. J., and Schaechter, M. (1988) *Cell* **54**, 127–135
- Russell, D. W., and Zinder, N. (1987) *Cell* **50**, 1071–1079
- Stancheva, I., Koller, T., and Sogo, J. M. (1999) *EMBO J.* **18**, 6542–6551
- Modrich, P. (1991) *Annu. Rev. Genet.* **25**, 229–253
- Marinus, M. G. (1985) *Mol. Gen. Genet.* **200**, 185–186
- Theisen, P. W., Grimwade, A. C., Leonard, J. A., Bogan, J. A., and Helmstetter, C. E. (1993) *Mol. Microbiol.* **10**, 575–584
- Roberts, J. M., Hoopes, B. C., McClure, W. R., and Kleckner, N. (1985) *Cell* **43**, 117–130
- Sternglanz, H., and Bugg, C. E. (1973) *Science* **182**, 833–834
- Engel, J. D., and Hoppel, P. H. (1978) *J. Biol. Chem.* **253**, 927–934
- Fazakerley, G. V., Téoule, R., Guy, A., Fritzsche, H., and Guschlbauer, W. (1985) *Biochemistry* **24**, 4540–4548
- Rinkel, L. J., van der Marel, G. A., van Boom, J. H., and Altona, C. (1987) *Eur. J. Biochem.* **163**, 275–286
- Fazakerley, G. V., Quignard, E., Téoule, R., Guy, A., and Guschlbauer, W. (1987) *Eur. J. Biochem.* **167**, 397–404
- Lingbeck, J., Kubine, M. G., Miller, J., Reid, B. R., Drobny, G. P., and Kennedy, M. A. (1996) *Biochemistry* **35**, 719–734
- Geek, H., and Freeman, R. (1991) *J. Magn. Reson.* **93**, 93–141
- Guéron, M., and Leroy, J.-L. (1995) *Methods Enzymol.* **261**, 383–413
- Borgias, B. A., and James, T. L. (1989) *Methods Enzymol.* **176**, 169–183
- Bae, S.-H., Cheong, H.-K., Lee, J.-H., Cheong, C., Kainosho, M., and Choi, B.-S. (2001) *Proc. Natl. Acad. Sci. U. S. A.* **98**, 10602–10607
- 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., Read, R. J., Rice, L. M., Simonson, T., and Warren, G. L. (1998) *Acta Crystallogr. Sect. D Biol. Crystallogr.* **54**, 905–921
- Koradi, R., Billeter, M., and Wüthrich, K. (1996) *J. Mol. Graph.* **14**, 51–55
- Lu, X.-J., Shakked, Z., and Olson, W. K. (2000) *J. Mol. Biol.* **300**, 819–840
- Lavery, R., and Skelenar, H. (1997) *CURVES 5.2, Helical Analysis of Irregular Nucleic Acids*, Biochimie Théorique, CNRS URA 77, Paris
- Engel, J. D., and Hoppel, P. H. (1974) *Biochemistry* **13**, 4143–4158
- Raszka, M. (1974) *Biochemistry* **13**, 4616–4622
- Michalczyk, R., and Russu, I. M. (1999) *Biophys. J.* **76**, 2679–2686
- Leonard, G., and Hunter, W. (1993) *J. Mol. Biol.* **234**, 198–208
- Várnai, P., Djuranovic, D., Lavery, R., and Hartmann, B. (2002) *Nucleic Acids Res.* **30**, 5398–5406

30. Chuprina, V. P., Lipanov, A. A., Fedoroff, O. Y., Kim, S.-G., Kintanar, A., and Reid, B. (1991) *Proc. Natl. Acad. Sci. U. S. A.* **88**, 9087–9091
31. Folta-Stogniew, E., and Russu, I. M. (1996) *Biochemistry* **35**, 8439–8449
32. Hud, N. V., Schultze, P., and Feigon, J. (1998) *J. Am. Chem. Soc.* **120**, 6403–6404
33. Guo, Q., Lu, M., and Kallenbach, N. R. (1995) *Biochemistry* **34**, 16359–16364
34. Guarné, A., Zhao, Q., Ghirlando, R., and Yang, W. (2002) *Nat. Struct. Biol.* **9**, 839–843
35. Kang, S., Lee, H., Han, J. S., and Hwang, D. S. (1999) *J. Biol. Chem.* **274**, 11463–11468
36. Brendler, T., and Austin, S. (1999) *EMBO J.* **18**, 2304–2310
37. Baikalov, I., Grzeskowiak, K., Yanagi, K., Quintana, J., and Dickerson, R. E. (1993) *J. Mol. Biol.* **231**, 768–784
38. Pastor, N., Weinstein, H., Jamison, E., and Brenowitz, M. (2000) *J. Mol. Biol.* **304**, 55–68
39. Guex, N., and Peitsch, M. C. (1997) *Electrophoresis* **18**, 2714–2723