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 N6 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 N6-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.

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 C5 and N4 positions of cytosine and the N6 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/GAATTC, or GAATTC/GAATTC, respectively, with A representing N6-methylated adenine). In Escherichia coli, most of the N6 positions of 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 hemimethylated 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 N6-methylated adenine base showed that the methyl group attached at the adenine N6 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 N6 would disrupt the secondary structure of the DNA double helix (11, 12). However, structural studies of double-stranded oligonucleotides revealed that N6-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 N6 hemimethylated DNA has a lower melting temperature and distinct dynamic characteristics represented by slow duplex-single strand exchange (13–16). Here, we investigate the structure and dynamics of two dodecamer DNA duplexes that contain either a single hemimethylated or unmethylated GATC site. Our study revealed that the N6-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
Structure and Dynamics of Hemimethylated GATC Site

EXPERIMENTAL PROCEDURES

Sample Preparation—All DNA oligonucleotides were synthesized on a solid support using an Applied Biosystems 391 DNA synthesizer using standard phosphoramidite chemistry. Adenine N°-methylated strands were synthesized by using N°-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 100 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) spectroscopy (NOESY) (τm = 240 ms) was carried out in 99.96% D2O at 30 °C, Two-dimensional NOESY (τm = 60, 120, and 240 ms), two-dimensional double quantum filter correlation spectroscopy (DQF-COSY), two-dimensional total correlation spectroscopy (TOCSY) (τm = 80 ms), 1H-13P heteronuclear correlation spectroscopy, and rotational NOE spectroscopy (ROESY) (τm = 240 ms) were conducted in 99.96% D2O at 30 °C. One-dimensional natural abundance 1H-13C HSQC spectra were obtained at 0, 6, 12, 30, 42, 48, 54, 60, and 66 °C with selective carbon pulses for the adenine N°-methyl group.

Base Pair Exchange Kinetics—Selective longitudinal relaxation times were measured at increasing concentrations of NH3 ranging from 0 to 0.6 mM 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 NH3+ Ions—DNA binding of NH3+ ions was probed by one-dimensional 15N-decoupled 1H-13C HSQC spectra, which were acquired with ~2 mM DNA, 10 mM sodium phosphate (pH 6.8), and 20 mM glycin. The 15N 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 (τm = 3, 4, and 5 ms) using a relaxation matrix analysis program, MARDIGRAS (19). The 6 dihedral angle was derived from the analysis of 13C-H1, HEN, double quantum filter correlation spectroscopy, χ was constrained to 220 ± 45° based on the medium-to-weak intra-residue H6/H8/H11 NOE. α and ζ were unconstrained, and other backbone dihedral angles were loosely constrained to the standard B-form (β (180 ± 30°), γ (60 ± 20°), and ε (230 ± 70°)), except that γ was unconstrained for the hemimethylated DNA. Scalar 1H-13C and dipolar 13C-1H couplings were derived from natural abundance 1H-13C HSQC experiments with and without Pf1 (~10 mg/ml) at 30 °C. Pf1 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−1Å2, and the dihedral angle force constant, which initially was 5, was scaled to 250 kcal mol−1radian−2 during cooling. The force constant for the residual dipolar coupling energy function was 3 kcal mol−1. For unmethylated DNA, 15 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 18 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 13C Resonances of N°-Methyl Group—In the H2O-NOESY (τm = 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/
temperatures ranging from 0 to 70 ppm (25). Another study of di-
agonal peak in a D2O rotational NOE spectroscopy spectrum
and that the cross-peak between them has the same sign as the
C6-N6 peak in H2O observed that the methyl resonance is a doublet
and that the amino proton resonance is a spin-coupled quartet
for the unmethylated DNA, and a few resonances have
line widths that are broader than the corresponding res-
ances in the absence of H2O. Therefore, we reasonably assigned
these two distinct resonances as the methyl resonances of the
two cis forms of N6-methylated adenine. In addition, previous
NMR studies performed at the mononucleotide level
support this resonance assignment. For example, an NMR
study of N6-methylated adenine in CDCl3 solvent showed that the
N6-methyl chemical shifts are 3.07 (cis) and 3.51 (trans)
ppm (25). Another study of an N6-methylated adenine mono-
mer in H2O 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 N6-Methylated Ade-
nine—A series of N6-methyl-selective natural abundance
1H-13C HSQC experiments performed at increasing temperatures
(0–66 °C) revealed that the trans-cis interconversion of the
N6-methylated adenine via C6-N6 bond rotation is quite slow
(Fig. 2E). From 0 to 12 °C the separation of the two peaks
remained at ~24 Hz, and the separation decreased to ~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 res-
nances in the absence of C6-N6 bond rotation is 24 Hz, the
simulated exchange kinetics estimates that the exchange pro-
cess should occur <10 times per second (s−1) 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 res-
nances 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−1 over
temperatures ranging from 0 to 70 °C (27). Because the C6-N6
bond length of N6-methylated adenine (1.342 Å) (11) is nearly
equivalent to that of a normal adenine C6-N6 bond length (1.335
Å), the slow exchange of the N6-methylated adenine should be
explained by structural reasons other than the partial double
bond character of the C6-N6 bond.

Overall Structures of Unmethylated and Hemimethylated
DNAs—For both unmethylated and hemimethylated DNA, NOE
connectivity and sugar puckers (C2-endo, 3JH12-H1’ > 8
Hz) typical of B-DNA were observed, and the 31P chemical
shifts (~3 ppm) were populated within a small range. How-
ever, one deviation from the standard B-DNA structure was
explained by structural reasons other than the partial double
bond character of the C6-N6 bond.

Characteristic Structural Identities of the Hemimethylated
GATC Site in Regard to the Backbone—For hemimethylated
DNA, the α/γ angles were correlated as gauche +/gauche − in
residues A4, A6, A16, A18, and C20. The γ 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 α/γ angles
gauche −/gauche +, a few metastable sub-states, including
α/‘γ-gauche +/gauche −, exist in DNA-protein complexes and
free DNA (29). This α/‘γ-gauche +/gauche − sub-state is accom-
panied by a B1 conformation, which is characterized by an α/γ;
trans/gauche − conformation with the (α/γ) ~90°. The
DNA backbones of both the unmethylated and hemimethylated DNA were B1 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 N⁰-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 ± 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 ± 0.9 Å for the cis and 10.9 ± 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 ± 1°) 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 ± 2° for the cis and 40 ± 1° for the trans), and the A6-T7/A18-T19 base step is under-twisted (23 ± 2° for the cis and 30 ± 1° for the 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 ²H,²⁵N HSQC of a ²⁵N-labeled ammonium ion (NH₄⁺) in which the ²H,²⁵N 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 ²H,²⁵N correlation peak was observed, both unmethylated and hemimethylated DNA showed a clear ²H,²⁵N correlation peak at least at the temperature at which we measured the base pair lifetimes (12°C) (Fig. 5C). Therefore,
hemimethylated 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$_2$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 $>5^\circ$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 $^1$H-$^15$N HSQC experiments with the $^{15}$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 $^1$H-$^15$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 (−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 (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).
forms. With regard to the crystal structure, the N\textsuperscript{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\textsuperscript{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°) 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\textsuperscript{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 are consistent with both the trans and cis conformations on an NMR time scale. However, the N\textsuperscript{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\textsuperscript{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°) 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\textsuperscript{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 are consistent with both the trans and cis conformations on an NMR time scale. However, the N\textsuperscript{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.
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., Schultz, 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. Guarne, 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
Structure and Dynamics of Hemimethylated GATC Sites: IMPLICATIONS FOR DNA-SeqA RECOGNITION
Sung-Hun Bae, Hae-Kap Cheong, Chaejoon Cheong, Sukhyun Kang, Deog Su Hwang and Byong-Seok Choi

J. Biol. Chem. 2003, 278:45987-45993.
doi: 10.1074/jbc.M306038200 originally published online August 27, 2003

Access the most updated version of this article at doi: 10.1074/jbc.M306038200

Alerts:
• When this article is cited
• When a correction for this article is posted

Click here to choose from all of JBC's e-mail alerts

This article cites 38 references, 6 of which can be accessed free at http://www.jbc.org/content/278/46/45987.full.html#ref-list-1