J|A|C|S
ARTICLES

Published on Web 12/09/2008

Structural and Dynamics Study of DNA Dodecamer Duplexes That Contain Un-, Hemi-, or Fully Methylated GATC Sites

Jongchul Bang,[†] Sung-Hun Bae,[†,§] Chin-Ju Park,[†,¶] Joon-Hwa Lee,[*,‡] and Byong-Seok Choi[*,†]

Department of Chemistry and National Creative Research Initiative Center, KAIST, 373-1, Guseong-dong, Yuseong-gu, Daejeon 305-701, Republic of Korea, and Department of Chemistry, RINS, and Environmental Biotechnology National Core Research Center, Gyeongsang National University, Jinju, Gyeongnam 660-701, Korea

Received May 22, 2008; E-mail: byongseok.choi@kaist.ac.kr; joonhwa@gnu.ac.kr

Abstract: Methylation of DNA plays a regulatory role in DNA metabolism. The *Escherichia coli* DNA adenine methyltransferase methylates the N6 positions of adenines in the sequence 5′-GATC-3′, which exists in the fully methylated state during most of the cell cycle. Just after DNA replication, however, the GATC sites transiently become hemimethylated, a condition that is indispensable for various cellular processes, such as negative modulation of replication initiation at *oriC* by SeqA. The lack of structural and dynamic information on DNA duplexes that contain fully methylated GATC sites makes it difficult to explain how hemimethylated GATC sites are recognized *in vivo* by proteins in a sea of fully methylated ones. Here, we used NMR spectroscopy to characterize the solution structure of a dodecamer DNA duplex that contained a fully methylated GATC site and the dynamics of the unmethylated, hemimethylated, and fully methylated GATC duplexes. Only the hemimethylated GATC duplex displays a unique major groove conformation, which is optimized for entrance into the cleft structure of SeqA. The apparent equilibrium constants for base-pair opening of the three differentially methylated GATC duplexes revealed that N6-methylation of the adenine residue affects the thermodynamics and kinetics of its own and neighboring base pairs. The equilibrium constants for base-pair opening of three GATC duplexes were determined using proton exchange catalyzed by TRIS. The two G·C base pairs of the hemimethylated GATC duplex displayed a faster base-pair opening rate and required less energy for the base-pair opening reaction than did those of the fully methylated one.

Introduction

Enzymatic methylation of DNA occurs abundantly in most living organisms and regulates a variety of cellular processes.[1] DNA methyltransferases commonly methylate the C5 and N4 positions of cytosine and the N6 position of adenine. The well-characterized DNA adenine (*dam*) methyltransferase of *Escherichia coli* (*E. coli*) is responsible for most methylation events that occur at the N6 positions of adenines within the 5′-GATC-3′ sequence.[2] The *E. coli* chromosome contains more than 19,000 GATC sites,[3] and 11 such sites are clustered in the 245 base pair (bp) region of the chromosome that contains the replication origin *oriC*. The fact that the number of GATC sites in the *oriC* region is greater than that predicted by statistical analysis suggests that these sites might participate in the regulation of DNA replication in *E. coli*.[4]

The *dam*-methylation status of the N6 position of adenines in GATC sites plays a key role as a signal for initiation of several cellular processes. For example, the *E. coli* SeqA protein inhibits the initiation of chromosomal replication at *oriC* on newly synthesized, hemimethylated GATC sites,[5] and the SeqA protein and a hemimethylated *oriC* are also associated with nucleoid segregation.[6,7] During DNA mismatch repair, transient hemimethylation of GATC sites after DNA replication signals to the MutH protein to cleave exclusively the unmethylated GATC DNA strand.[8] These phenomena are thought to be caused by the structural and dynamic effects of *dam*-methylation on DNA. Methylation of adenine residues is known to affect the stability of nucleic acid structures[9] and to change DNA curvature.[10]

The first three-dimensional (3-D) structure of an N6-methylated adenine residue was determined 30 years ago by X-ray crystallography.[11] Since then, various studies have focused on

[†] Korea Advanced Institute of Science and Technology.
[‡] Gyeongsang National University.
[§] Present address: Department of Molecular Biology, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, CA 92037.
[¶] Present address: Department of Biochemistry, Molecular Biology and Biophysics, University of Minnesota, Minneapolis, MN 55455.

(1) Jeltsch, A. *ChemBioChem* **2002**, *3*, 274–93.
(2) Geier, G. E.; Modrich, P. *J. Biol. Chem.* **1979**, *254*, 1408–13.
(3) Henaut, A.; Rouxel, T.; Gleizes, A.; Moszer, I.; Danchin, A. *J. Mol. Biol.* **1996**, *257*, 574–85.
(4) Zyskind, J. W.; Smith, D. W. *Cell* **1986**, *46*, 489–90.
(5) Lu, M.; Campbell, J. L.; Boye, E.; Kleckner, N. *Cell* **1994**, *77*, 413–26.
(6) Bach, T.; Krekling, M. A.; Skarstad, K. *EMBO J.* **2003**, *22*, 315–23.
(7) Ogden, G. B.; Pratt, M. J.; Schaechter, M. *Cell* **1988**, *54*, 127–35.
(8) Modrich, P. *Annu. Rev. Genet.* **1991**, *25*, 229–53.
(9) Engel, J. D.; von Hippel, P. H. *J. Biol. Chem.* **1978**, *253*, 927–34.
(10) Diekmann, S. *EMBO J.* **1987**, *6*, 4213–7.
(11) Sternglanz, H.; Bugg, C. E. *Science* **1973**, *182*, 833–4.

 10.1021/ja8038272 CCC: $40.75

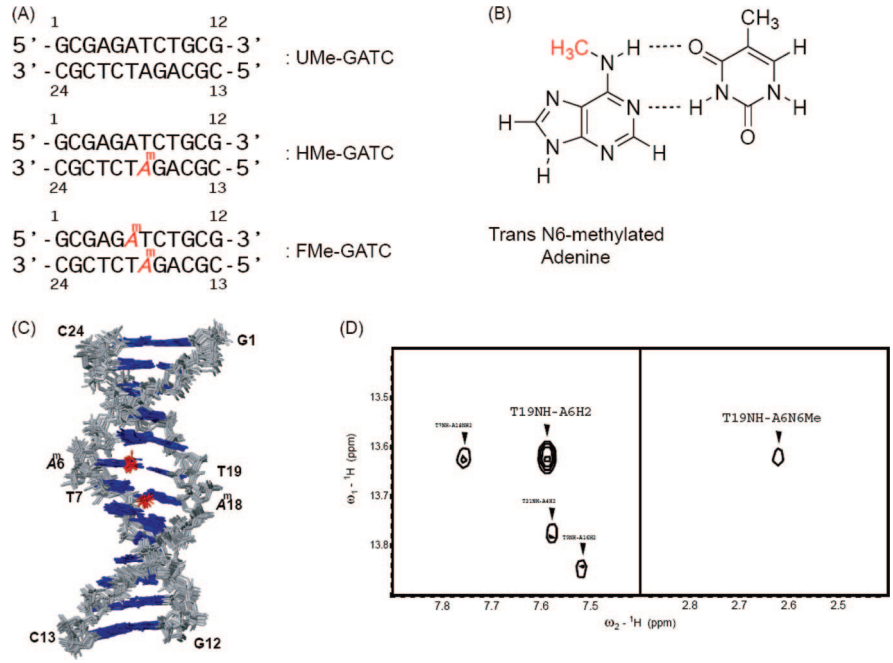

Figure 1. (A) Sequence and numbering of the dodecamer DNA duplexes that contain (from top to bottom) unmethylated, hemimethylated, and fully methylated GATC sites (referred to as the UMe-GATC, HMe-GATC, and FMe-GATC duplexes, respectively). A^m indicates an N6-methylated adenine. (B) The Watson−Crick base pairing between a *trans*-N6-methylated adenine residue and its complementary thymine residue. (C) Superimposed overall structure (12 structures) of the FMe-GATC duplex. Red bonds in the overall structure represent the N6-methyl moieties of the mA6 and mA18 residues. (D) Section of the 2-D Watergate NOESY spectrum of the FMe-GATC duplex. Cross peaks between T19-H3 and A6-H2 (left) and T19-H3 and A6-N6Me (right) are shown.

hemimethylated GATC sites and their related proteins.[12−18] It has been shown that the N6-methylation of adenine as well as the presence of four bases in the hemimethylated GATC site are required for SeqA binding.[14] Interestingly, the crystal structure of SeqA bound to hemimethylated DNA showed that the G·C base pair 5′ to the unmethylated adenine did not form any important interaction with the protein.[16] Comparative study of both hemimethylated and unmethylated GATC sites showed that the hemimethylated GATC site exhibited an unusual metastable backbone structure and a narrower major groove width compared to unmethylated GATC.[19] This study suggests that these unique structural features are important for the specific recognition of hemimethylated GATC sites by the SeqA protein. The hemimethylated GATC status, however, exists only transiently during the DNA replication process, and the dominant structure is a fully methylated GATC site. Consequently, it is necessary to investigate the properties of fully methylated GATC sites in order to understand fully the correlation between *dam*-methylation and SeqA−DNA interaction.

In addition, specific binding of SeqA to hemimethylated DNA cannot be explained by structural information alone. Thus, it is crucial that researchers investigate the dynamic features of the GATC site in order to elucidate the mechanism of specific recognition by SeqA. NMR is one of the most useful methods for dynamic studies of nucleic acids.[20−27]

Thus, we characterized the solution structure of a dodecamer DNA duplex that contains a single, fully methylated GATC site (referred to as the FMe-GATC duplex, Figure 1A) and compared its unique structural features with previously reported structures of unmethylated (UMe-GATC) and hemimethylated (HMe-GATC) duplexes.[19] We also investigated the kinetics and thermodynamics of base-pair openings of the three GATC-containing DNA duplexes (FMe-, HMe-, and UMe-GATC) using TRIS as a base catalyst. From our structural study, we found that double N6-methylation at the GATC site restored the narrowed major groove conformation observed in the HMe-GATC duplex to the groove conformation observed in the UMe-GATC duplex. We hypothesized that peculiar structure of the major groove around the central GATC site could be important for the SeqA-DNA interaction. We also observed changes in the base-pair lifetimes and apparent equilibrium constants for base-pair opening in concert with the N6-methylation of adenines in the central GATC sites. The presence and absence of additional N6-methylated adenine residues in the duplex

(12) Guo, Q.; Lu, M.; Kallenbach, N. R. *Biochemistry* **1995**, *34*, 16359–64.
(13) Lingbeck, J.; Kubinec, M. G.; Miller, J.; Reid, B. R.; Drobny, G. P.; Kennedy, M. A. *Biochemistry* **1996**, *35*, 719–34.
(14) Brendler, T.; Austin, S. *EMBO J.* **1999**, *18*, 2304–2310.
(15) Kang, S.; Lee, H.; Han, J. S.; Hwang, D. S. *J. Biol. Chem.* **1999**, *274*, 11463–11468.
(16) Guarne, A.; Zhao, Q.; Ghirlando, R.; Yang, W. *Nat. Struct. Biol.* **2002**, *9*, 839–43.
(17) Fujikawa, N.; Kurumizaka, H.; Nureki, O.; Tanaka, Y.; Yamazoe, M.; Hiraga, S.; Yokoyama, S. *Nucleic Acids Res.* **2004**, *32*, 82–92.
(18) Lee, J. Y.; Chang, J.; Joseph, N.; Ghirlando, R.; Rao, D. N.; Yang, W. *Mol. Cell* **2005**, *20*, 155–66.
(19) Bae, S. H.; Cheong, H. K.; Cheong, C.; Kang, S.; Hwang, D. S.; Choi, B. S. *J. Biol. Chem.* **2003**, *278*, 45987–93.
(20) Lee, J. H.; Pardi, A. *Nucleic Acids Res.* **2007**, *35*, 2965–74.
(21) Varnai, P.; Canalia, M.; Leroy, J. L. *J. Am. Chem. Soc.* **2004**, *126*, 14659–67.
(22) Snoussi, K.; Leroy, J. L. *Biochemistry* **2002**, *41*, 12467–74.
(23) Warmlander, S.; Sandstrom, K.; Leijon, M.; Graslund, A. *Biochemistry* **2003**, *42*, 12589–95.
(24) Warmlander, S.; Sen, A.; Leijon, M. *Biochemistry* **2000**, *39*, 607–15.
(25) Cain, R. J.; Glick, G. D. *Biochemistry* **1998**, *37*, 1456–64.
(26) Bhattacharya, P. K.; Cha, J.; Barton, J. K. *Nucleic Acids Res.* **2002**, *30*, 4740–50.
(27) Dornberger, U.; Leijon, M.; Fritzsche, H. *J. Biol. Chem.* **1999**, *274*, 6957–62.

induced considerable differences between the dynamic properties of the HMe-GATC duplex and the FMe-GATC duplex. We suggest that the dynamic features identified in this study are important for deciphering how hemimethylated GATC sites are recognized by proteins in a background of many more fully methylated ones *in vivo.*

Experimental Procedures

Sample Preparation. All DNA oligonucleotides were purchased from Genotech Co., LTD (Daejeon, Korea). The oligonucleotides were desalted using a Sephadex G-25 gel filtration column, and 0.5 to 1 mM DNA duplex samples were formed by combining molar equivalents of the two strands in a 90% H_2O/10% D_2O NMR buffer [10 mM sodium phosphate (pH 6.8) and 100 mM NaCl]. For the TRIS-catalyzed experiments, the buffer was changed to 10 mM TRIS (pH 8.49 at 24.2 °C) and 100 mM NaCl. The TRIS concentration was increased from 10 mM to 196 mM by successive additions of a 500 mM TRIS stock solution. The pH of the sample dissolved in TRIS buffer was calculated using the relation $\Delta pK_a = -0.031 \times \Delta T$.[28]

Imino Proton Exchange Theory. The formalism of catalyzed proton exchange has been extensively described.[29,30] It is assumed that the imino proton exchange from a base pair consisted of a two-step process requiring base-pair opening followed by proton transfer to a base catalyst. The base catalyst contribution to the rate constant for imino proton exchange, $k_{ex,B}$, is given by eq 1,[20]

$$k_{ex,B} = \frac{k_{op}k_{tr}}{k_{cl} + k_{tr}} = \frac{\alpha K_{op}k_i[B]}{1 + \alpha K_{op}k_i[B]/k_{op}} \quad (1)$$

where k_{tr} is the imino proton transfer rate constant from the mononucleotide, k_{op} and k_{cl} are the rate constants for base-pair opening and closing, respectively, k_i is the rate constant for imino proton transfer, [B] is the base catalyst concentration, α is the accessibility factor, which is equal to one in the case of a fully accessible imino proton, and K_{op} ($=k_{op}/k_{cl}$) is the equilibrium constant for base-pair opening.

NMR Experiments. NMR experiments were performed on a Varian Inova 600 MHz spectrometer. All 2-D data were processed with the program NMRPipe[31] and analyzed with the program Sparky,[32] and 1-D data were processed with the program FELIX (Accelrys). The 2-D Watergate-NOESY ($\tau_m = 240$ ms) was carried out in 90% H2O/10% D2O at 15 and 35 °C. The 2-D NOESY (τm = 60, 150 and 250 ms), 2-D COSY, and 2-D TOCSY (τm = 60 ms) were conducted in 100% D2O at 14 °C. Scalar $^1J_{CH}$ and dipolar $^1D_{CH}$ couplings of the FMe-GATC duplex were derived from natural abundance sensitivity enhancement $^1H-^{13}C$ HSQC experiments carried out at 30 °C with and without Pf1 (~15 mg/mL). The data were collected on a Varian Inova 900 MHz spectrometer equipped with an HCN triple resonance probe (KIST, Seoul). Pf1 filamentous bacteriophage was purchased from ASLA, Ltd. Semiselective inversion recovery 1-D NMR experiments were used to determine the apparent longitudinal relaxation rate constants, R_{1a} ($= 1/T_{1a}$), of the imino protons and water, R_{1w}, as described previously.[20] The water magnetization transfer experiments were used to determine the hydrogen exchange rates of the imino protons as described previously.[20] The exchange rate constants (k_{ex}) for imino protons were determined by curve fitting of the difference data, of peak intensities between control and time t, $[I_0 - I(t)]/I_0$, to:

$$\frac{I_0 - I(t)}{I_0} = 2\frac{k_{ex}}{(R_{1w} - R_{1a})}\left(e^{-R_{1a}t} - e^{-R_{1w}t}\right) \quad (2)$$

where R_{1a} and R_{1w} were previously measured by inversion recovery.[20]

(28) Mohan, C. *Buffers: A Guide for the Preparation and Use of Buffers in Biological Systems*; CalBiochem: Darmstadt, Germany, 2003.

Table 1. Structure Determination Statistics for the FMe-GATC Duplex

	FMe-GATC duplex
total number of NOE distance restraints	352
intraresidue distances	181
sequential residue distances	139
interstrand distances	32
dihedral restraints (β, γ, δ, ϵ, and χ)	114
base planarity restraints	12
residual dipolar coupling restraints	22
total number of restraints	500
pairwise rmsd for all heavy atoms (Å)	0.96 ± 0.42
average NOE violations (Å)	0 (>0.5 Å)
average dihedral angle violations (deg)	0 (>5)
Rmsd from Covalent Geometry	
bond lengths (Å)	0.009
angles (deg)	1.145
impropers (deg)	1.986

Structure Calculations. The distance constraints of the FMe-GATC duplex were derived from the integrated NOE volumes and three assumed isotropic correlation times ($\tau_c = 3$, 4, and 5 ns) using the relaxation matrix analysis program, MARDIGRAS.[33] The δ dihedral angle was derived from the analysis of $^3J_{H1'-H2'}$ in regular 2-D COSY.[34] The χ dihedral angle was constrained to 220 ± 45° on the basis of medium-to-weak intraresidue H6/H8−H1′ NOE. The α and ζ dihedral angles were unconstrained, and other backbone dihedral angles were loosely constrained to the standard B-form [β (180 ± 45°), γ (60 ± 30°), and ε (230 ± 70°)]. These three dihedral angles were deduced by analyzing the corresponding regions of NMR spectra according to a previously describe method.[35] All structure calculations were performed by using XPLOR-NIH[36] with restrained molecular dynamics. 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 150 ps of torsion angle dynamics cooling to 0 K. The final structures were generated after 20,000 cycles of energy minimization. The distance force constant was 40 kcal/mol·Å^2, and the dihedral angle force constant, which initially was 5, was scaled to 250 kcal/mol·rad^2 during cooling. Twelve of 100 trial structures were converged for the FMe-GATC duplex. The FMe-GATC duplex structure was refined with RDCs in addition to NOE and torsion angle restraints. Alignment tensor analysis of the observed RDC was performed as described previously.[37,38] The values $D_a = -21.0$ Hz and $R = 0.45$ were used in the calculation. The refinement step consists of an initial equilibration stage where the dipolar coupling force constants were increased from 0.01 to 5.0 kcal/mol·Hz^2 over 50 cycles, corresponding to a 15-ps molecular dynamics run. This was followed by a slow cooling from 1000 to 300 K in 15 cycles of molecular dynamics. During this cooling stage, the force constant for dipolar coupling was 5.0 kcal/mol·Hz^2. The final structures were generated after 2000 cycles of energy minimization.[37] Twelve structures were converged for the FMe-GATC duplex (Table 1). The final structures were analyzed with MOLMOL,[39] and Curves 5.3[41] software.

Determination of the Rate Constants for Base-Pair Opening and Closing, and the Apparent Equilibrium Constant for Base-Pair Opening. The apparent relaxation rate constant for an imino proton, R_{1a}, can be represented by:

(29) Gueron, M.; Leroy, J. L. *Methods Enzymol.* **1995**, *261*, 383–413.
(30) Leroy, J. L.; Bolo, N.; Figueroa, N.; Plateau, P.; Gueron, M. *J. Biomol. Struct. Dyn.* **1985**, *2*, 915–39.
(31) Delaglio, F.; Grzesiek, S.; Vuister, G. W.; Zhu, G.; Pfeifer, J.; Bax, A. *J. Biomol. NMR* **1995**, *6*, 277–93.
(32) Goddard, T. D.; Kneller, D. G. SPARKY; University of California, San Franscio, 2003.
(33) Borgias, B. A.; James, T. L. *Methods Enzymol.* **1989**, *176*, 169–83.
(34) Delaglio, F.; Wu, Z.; Bax, A. *J. Magn. Reson.* **2001**, *149*, 276–81.

$$R_{1a} = R_1 + k_{ex} = R_1 + k_{AAC} + k_{ex,B} \quad (3)$$

where R_1 is the relaxation rate constant of the imino protons, and k_{AAC} and $k_{ex,B}$ are the contribution of the intrinsic base and the added catalyst to imino 3DNA,[40] proton exchange, respectively. This equation is valid in the assumption that the R_1 value of the imino proton is not affected by addition of the base catalyst. Substituting eq 1 into eq 3 yields the following equation:

$$R_{1a} = R_1 + k_{AAC} + \frac{k_{op}k_i[B]}{k_{cl}/\alpha + k_i[B]} \quad (4)$$

The constants k_{cl}/α and k_{op} can be determined by curve fitting R_{1a} of the imino protons as a function of the concentration of the base catalyst with eq 4, and αK_{op} can be calculated by $\alpha k_{op}/k_{cl}$.

Results

Structure Determination and Overall Structure of the FMe-GATC Duplex. In the Watergate-NOESY spectrum of the FMe-GATC duplex, we observed only a single set of NOE peaks and strong cross-peaks between the H2 of the N6-methylated adenine (mA) and the imino proton of the complementary thymine residue, which are typical of Watson−Crick A·T base pairs. The intensities of these cross-peaks were stronger than were the intensities of the cross-peaks between the N6-methyl group and imino proton of the complementary thymine residue (Figure 1D). In addition, the base-pair opening kinetics is consistent with normal Watson−Crick base pairing of the target DNA substrates. Therefore, for our structural studies, we used only one isomer of the FMe-GATC duplexes (mA6:Trans/mA18:Trans) in which both N6-methylated adenine residues were in the trans configuration, whereas two isomeric structures were determined for the HMe-GATC duplex in a previous study.[19]

For the structure calculation, the backbone dihedral angles α and ζ were unconstrained, but the other angles were constrained in the range of standard B-form DNA. All residues in the FMe-GATC duplex showed H1′-H2′ scalar couplings larger than 8 to 9 Hz, meaning that they have the *C2′-endo* sugar puckers typical of B-form DNA. Three dihedral angles β, γ, and ϵ were confirmed using previously reported method.[35] Figure 2 shows a comparison of the major grove width determined here for the FMe-GATC duplex and those for the UMe- and HMe-GATC duplexes that were reported previously. The major groove width at the GATC site in the HMe-GATC duplex was reduced dramatically compared to the UMe-GATC duplex (Figure 2).[19] However, the additional N6-methylated adenine in the GATC site restored this unusual major groove conformation so that it resembled that of the UMe-GATC duplex (Figure 2).

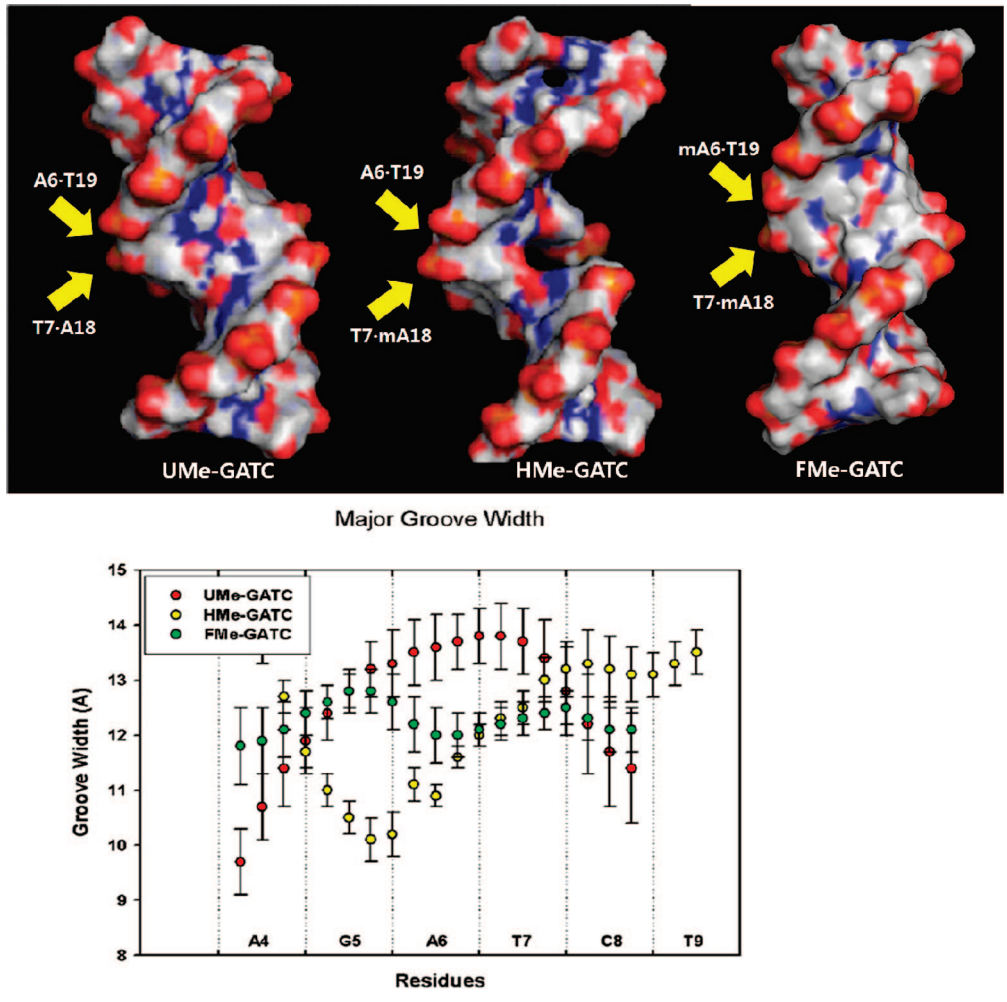

Figure 2. Surface representations (top) and major groove widths (graph, bottom) of the UMe-, HMe-, and FMe-GATC duplexes. Surface models were prepared with the program Pymol.[42] Groove widths were calculated with Curves 5.3[41] software. The error bars represent the standard deviations.

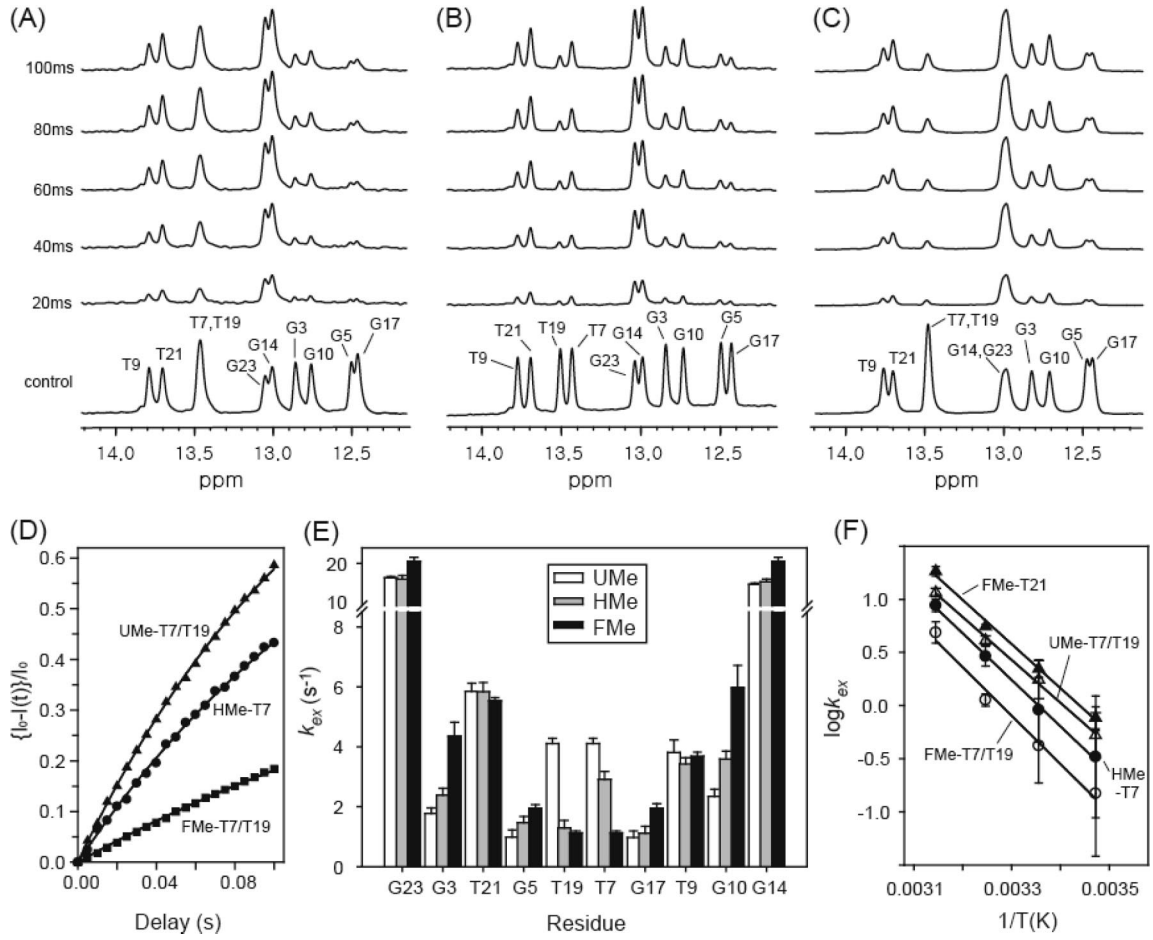

Figure 3. Difference spectra between the control (no selective inversion for water) and exchange experiments after selective inversion in water magnetization transfer experiments of imino protons in the (A) UMe-, (B) HMe-, and (C) FMe-GATC duplexes conducted at 35 °C. The delay time after selective water inversion are shown on the left of each spectrum. 2-D Watergate NOESY spectra were used to make the resonance assignments for the imino protons. Residue numbers are indicated in the control spectra. (D) Relative peak intensities, $[I(t) - I_0]/I_0$, for imino protons of the central A·T base pairs as a function of delay time. The solid lines are the best fits to eq 2. (E) The rate constants, k_{ex}, for exchanges of the imino protons of the UMe-, HMe-, and FMe-GATC duplexes at 35 °C. The error bars represent fitting errors. (F) Logarithm scale of the k_{ex} values for the imino protons of thymine residues T7 and T19 in the central GATC sites versus the inverse of the temperature.

Resonance Assignment and Hydrogen Exchange Rate Measurement for the Imino Protons in the GATC Duplexes. A previous kinetics study of the UMe-GATC and HMe-GATC duplexes revealed that the base pair lifetime of the HMe-GATC duplex is longer than that of the UMe-GATC duplex in the central A·T base pairs.[19] In order to investigate in detail the dynamic features of the base pairs in the UMe-, HMe-, and FMe-GATC duplexes, we performed water magnetization transfer experiments and base-catalyzed hydrogen exchange measurements of the imino protons of the three GATC duplexes. All imino protons were assigned based on the 2-D Watergate-NOESY spectra of each duplex. Most of the imino proton resonances were distinct and well separated, except for those of the T7 and T19 residues in the central GATC site in the FMe- and UMe-GATC duplexes and the G14 and G23 residues in the FMe-GATC duplex.

By performing semiselective inversion recovery experiments on the imino proton resonances, we determined the apparent longitudinal relaxation rate constants, R_{1a}, for the imino protons of the UMe-, HMe-, and FMe-GATC duplexes at 5, 15, 25, 35 and 45 °C. Except for the 5 °C results, all R_1 data for the imino protons fit well to a single exponential relaxation function (data not shown). Thus, hydrogen exchange rates for the imino protons were determined from the water magnetization transfer experiments performed at 15, 25, 35, and 45 °C.

Figure 3 (A−C) shows the 1-D difference spectra of the water magnetization transfer experiments for the UMe-, HMe-, and FMe-GATC duplexes at 35 °C. Figure 3D shows the result of curve fitting the peak intensities of the T7 imino protons in three duplexes. The k_{ex} values of all imino protons in the three duplexes, except those in the terminal base pairs, are shown in Figure 3E. The N6-methyl modification at mA18 in the HMe-GATC duplex reduced the k_{ex} values of the imino protons of both T7 (2.9 s^{-1}) and T19 (1.3 s^{-1}) in the central GATC site (Figure 3E). The additional N6-methyl modification at mA6 in the FMe-GATC duplex yielded k_{ex} values for base pairs

(35) Kim, S. G.; Lin, L. J.; Reid, B. R. *Biochemistry* **1992**, *31*, 3564–74.
(36) Schwieters, C. D.; Kuszewski, J. J.; Tjandra, N.; Clore, G. M. *J. Magn. Reson.* **2003**, *160*, 65–73.
(37) Tjandra, N.; Tate, S.; Ono, A.; Kainosho, M.; Bax, A. *J. Am. Chem. Soc.* **2000**, *122*, 6190–6200.
(38) Zweckstetter, M.; Bax, A. *J. Am. Chem. Soc.* **2000**, *122*, 3791–3792.
(39) Koradi, R.; Billeter, M.; Wuthrich, K. *J. Mol. Graphics* **1996**, *14*, 51–5.
(40) Lu, X. J.; Olson, W. K. *Nucleic Acids Res.* **2003**, *31*, 5108–21.
(41) Lavery, R.; Sklenar, H. *J. Biomol. Struct. Dyn.* **1988**, *6*, 63–91.
(42) DeLano, W. L. *The PyMOL Molecular Graphics System*; 2002; http://www.pymol.org.

Table 2. Parameters for Base Pair Stability and Dynamics of the GATC-Containing DNA Duplexes at 35 °C[a]

base pair	duplex	αK_{op} (10^{-6})	τ_0 (=$1/k_{op}$) (ms)	$\Delta G^0_{base\ pair}$[b] (kcal/mol)	$\Delta\Delta G^0_{base\ pair}$[c] (kcal/mol)	$\Delta\Delta G^{\ddagger}_{opening}$[d] (kcal/mol)	$\Delta\Delta G^{\ddagger}_{closing}$[e] (kcal/mol)
G3·C22	UMe	5.16 ± 0.06	0.3 ± 0.3	−7.45 ± 0.01	−0.01 ± 0.01	−1.12 ± 0.65	−1.13 ± 0.65
	HMe	5.29 ± 0.06	2.0 ± 0.4	−7.44 ± 0.01	–	–	–
	FMe	5.98 ± 0.13	3.9 ± 0.5	−7.36 ± 0.01	0.07 ± 0.02	0.40 ± 0.15	0.47 ± 0.15
G5·C20	UMe	0.73 ± 0.01	9.7 ± 1.8	−8.65 ± 0.01	−0.03 ± 0.02	−0.56 ± 0.15	−0.59 ± 0.15
	HMe	0.76 ± 0.02	24.4 ± 4.3	−8.62 ± 0.02	–	–	–
	FMe	0.59 ± 0.02	64.1 ± 7.2	−8.78 ± 0.02	−0.16 ± 0.03	0.59 ± 0.13	0.43 ± 0.13
A6·T19	UMe	41.93 ± 0.67	1.4 ± 0.2	−6.17 ± 0.01	0.54 ± 0.01	−0.36± 0.14	0.18 ± 0.14
	HMe	17.47 ± 0.28	2.4 ± 0.4	−6.71 ± 0.01	–	–	–
	FMe	8.33 ± 0.12	2.5 ± 0.5	−7.16 ± 0.01	−0.45 ± 0.01	0.01 ± 0.17	−0.44 ± 0.17
T7·A18	UMe	41.93 ± 0.67	1.4 ± 0.2	−6.17 ± 0.01	0.49 ± 0.01	−0.36 ± 0.13	−0.36 ± 0.13
	HMe	18.69 ± 0.25	2.4 ± 0.4	−6.66 ± 0.01	–	–	–
	FMe	8.33 ± 0.12	2.5 ± 0.5	−7.16 ± 0.01	−0.50 ± 0.01	0.01 ± 0.26	−0.49 ± 0.26
C8·G17	UMe	0.79 ± 0.01	8.1 ± 1.4	−8.60 ± 0.01	0.29 ± 0.02	−0.03 ± 0.43	0.26 ± 0.43
	HMe	0.49 ± 0.01	8.5 ± 5.7	−8.89 ± 0.02	–	–	–
	FMe	0.64 ± 0.02	64.8 ± 4.6	−8.72 ± 0.02	0.17 ± 0.02	1.25 ± 0.41	1.42 ± 0.41
G10·C15	UMe	8.55 ± 0.15	1.3 ± 0.3	−7.14 ± 0.01	−0.14 ± 0.01	−0.56 ± 0.16	−0.70 ± 0.16
	HMe	10.77 ± 0.15	3.3 ± 0.3	−7.00 ± 0.01	–	–	–
	FMe	12.28 ± 0.39	5.6 ± 0.5	−6.92 ± 0.02	0.08 ± 0.02	0.32 ± 0.08	0.40 ± 0.08
GATC site[f]	UMe			−29.59 ± 0.02	1.29 ± 0.04	−1.31 ± 0.49	−0.02 ± 0.49
	HMe			−30.88 ± 0.03	–	–	–
	FMe			−31.82 ± 0.03	−0.94 ± 0.04	1.86 ± 0.53	0.92 ± 0.53

[a] 35 °C, 100 mM NaCl, pH 8.5, 10−100 mM TRIS-d_{11} [b] $\Delta G^0_{base\ pair} = -\Delta G^0_{opening} = RT\ \ln(\alpha K_{op})$, T = 308 K. [c] $\Delta\Delta G^0_{base\ pair} = \Delta G^0_{base\ pair} - \Delta G^0_{base\ pair}{}^{HMe} = RT\ \ln(\alpha K_{op}/\alpha K_{op}{}^{HMe})$, T = 308 K. [d] $\Delta\Delta G^{\ddagger}_{opening} = \Delta G^{\ddagger}_{opening} - \Delta G^{\ddagger}_{opening}{}^{HMe} = -RT\ \ln(k_{op}/k_{op}{}^{HMe}) = RT\ \ln(\tau_0/\tau_0{}^{HMe})$, T = 308 K. [e] $\Delta\Delta G^{\ddagger}_{closing} = \Delta\Delta G^{\ddagger}_{opening} - \Delta\Delta G^{\ddagger}_{opening} = \Delta\Delta G^{\ddagger}_{opening} + \Delta\Delta G^0_{base\ pair}$. [f] Summation of data for the G5·C20, A6·T19, T7·A18, and C8·G17 base pairs.

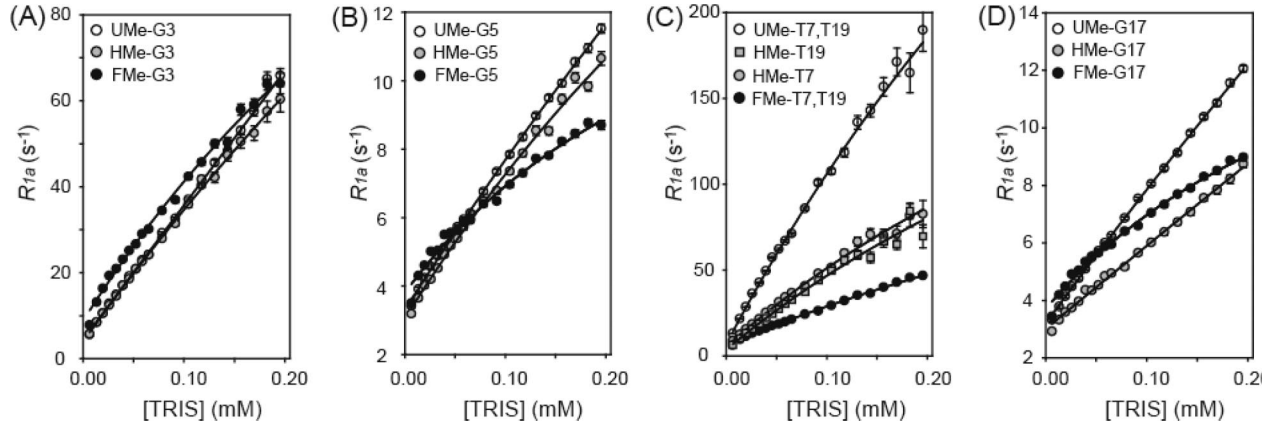

Figure 4. TRIS-catalyzed exchange carried out with the GATC duplexes. Apparent spin−lattice R_1 relaxation rate constants (R_{1a}) for the (A) G3, (B) G5, (C) T7/T19, and (D) G17 imino protons as a function of the TRIS concentrations are shown. The solid lines are the best fits to eq 4, and the error bars represent the fitting errors during determination of R_{1a} from inversion recovery data.

mA6·T19 and T7·mA18 that were 4-fold less than those of the corresponding base pairs in the UMe-GATC duplex. However, base pairs involving the G5 and G17 imino protons, which lie adjacent to the N6-methylated adenine residues, showed k_{ex} values that were slightly larger than those of the UMe-GATC duplex.

Figure 3F shows the temperature-dependence of the k_{ex} values of the imino protons for the UMe-, HMe-, and FMe-GATC duplexes. The linear correlation between log k_{ex} and $1/T$ indicates the Arrhenius equation, and the slopes of these lines yield the activation energy of the hydrogen exchange process.

Base-Pair Opening Stability and Kinetics in the Presence of the TRIS Base Catalyst. The differences in the k_{ex} values among the three GATC duplexes studied herein were the largest at 45 °C, but this high temperature rendered the duplexes too flexible to measure the kinetic parameters. Thus the base-catalyzed exchange experiments with the GATC duplexes were performed at 35 °C, which is a physiological condition for *E. coli*. TRIS base is a more useful catalyst for hydrogen exchange studies at high temperature than are other strong base catalysts, such as ammonia, which causes severe line-broadening and rapid exchange of the imino protons in the GATC duplexes. At NH_3 concentrations >50 mM, hydrogen exchange rates could be measured only for the G5 and G17 imino protons of the three GATC duplexes (data not shown). The αK_{op} and τ_0 (=$1/k_{op}$) values for most Watson−Crick base pairs in the three GATC duplexes were determined from analysis of the R_{1a} data as a function of the TRIS base concentration using eq 4 (Table 2 and Figure 4). The G3·C22 base pair, which is three base pairs apart from the N6-methylated adenine (A6), has similar αK_{op} values ($\sim 5 \times 10^{-6}$) for all three duplexes (Table 2), consistent with the similar slopes seen in Figure 4A. Similar results were observed for the G10·C15 base pair ($\alpha K_{op} = (9-13) \times 10^{-6}$) located in the other direction of the GATC site (Table 2). The τ_0 (=$1/k_{op}$) values of these two base pairs in all three duplexes were ≤6 ms (Table 2). The αK_{op} and τ_0 values for the A4·T21 and T9·A16 base pairs in all three GATC duplexes could not be determined because of severe line-broadening of the T21 and T9 imino resonances (data not shown).

In the HMe-GATC duplex, the N6-methylation at A18 (mA18) gives rise to αK_{op} values for the A6·T19 and T7·A18 base pairs that are half of those of the corresponding base pairs in the UMe-GATC duplex (Table 2). In the FMe-GATC duplex, the mA6·T19 and T7·mA18 base pairs, which both contain N6-methylated adenines, have αK_{op} values that are one-half and one-quarter of those of the corresponding base pairs in the HMe- and UMe-GATC duplexes, respectively (Table 2 and Figure 4C), indicating that the effects of the two N6-methylations are additive. The τ_0 values for the A6·T19 and T7·A18 base pairs in the all three duplexes are less than 5 ms (Table 2) and thus cannot be determined precisely.

The N6-methylation modification at the A·T base pairs also affected the dynamics of the C8·G17 base pair, which resides next to the N6-methylated mA18 residue (Figure 4D). The αK_{op} value of the C8·G17 base pair in the UMe-GATC duplex was slightly larger (1.5-fold) than that of the corresponding base pair in the HMe-GATC duplex, but this N6-methylation had little effect on the stability of the C20·G5 base pair, which is 2 base pairs apart from mA18 (Table 2). These results demonstrate that asymmetric N6-methylation at the central GATC site significantly stabilizes its own and the two neighboring base pairs and has a small effect on the dynamic properties of the one remaining base pair of the GATC site. The G5·C20 and C8·G17 base pairs of the FMe-GATC duplex have similar αK_{op} values to those of the other two duplexes. Surprisingly, the base-pair lifetime (τ_0) of the C8·G17 base pair in the FMe-GATC duplex was 8-fold longer than that of the corresponding base pair in the HMe-GATC duplex (Figure 4 and Table 2). A similar result was observed for the G5·C20 base pair (Figure 4 and Table 2). These results demonstrate that N6-methylations at both adenines in the central GATC site of the DNA duplex give rise to both slow base-pair opening and slow base-pair closing for the two adjacent G·C base pairs, but result in little change in G·C base pair stability.

Discussion

The *dam*-methylation status of the N6 position of adenine residues in GATC sites is known to be crucial for several biological processes and the specificity of protein−DNA interactions. Two possible mechanisms have been proposed to explain how these GATC-binding proteins function; these include (i) direct recognition of methyl groups in the DNA binding site and/or (ii) recognition of methylation-induced structural and dynamic changes in the DNA double helix. The former takes into account the different affinities of GATC-binding proteins for unmethylated and hemimethylated GATC sites; binding studies have shown that these proteins favor hemimethylated sites, owing to the presence of the N6-methyl group. Furthermore, the cocrystal structure of SeqA bound to hemimethylated DNA revealed that the N6-methyl group on the A residue of the DNA binding site makes van der Waals interactions with the T151 and R152 residues of the SeqA protein.[16] These features, however, cannot explain the observed differences in the affinities of GATC-binding proteins for hemimethylated and fully methylated GATC sites, as both sites possess N6-methyl groups. Therefore, to decipher the mechanism behind these differences in binding affinities, we focused on the structural and dynamic changes in the DNA double helix that are induced by N6-methylation.

An alanine scanning mutagenesis study revealed that certain arginine residues (R116, R118, and R155) are very important for SeqA−DNA complex formation.[17] A cocrystal structure of the SeqA−DNA complex showed that these arginine residues constitute the cleft structure of the SeqA protein, where the major groove side of the DNA duplex is inserted.[16,17] Structural comparisons between the hemimethylated and unmethylated GATC sites revealed that the difference in major groove width is 1−3 Å (Figure 2).[19] In the case of a fully methylated GATC duplex, the additional N6-methylation at the GATC site changes the narrowed major groove structure induced by hemimethylation to the normal major groove conformation. Thus, to avoid a steric clash between the cleft structure of SeqA and the major groove of a fully methylated GATC DNA duplex, a conformational change in the backbone of the DNA duplex must occur during formation of the SeqA−DNA complex. It has been suggested that this conformational change is responsible for the reduced binding affinity of the SeqA protein for a fully methylated GATC site, compared to a hemimethylated one.

Dynamic features of the base pairs in the GATC DNA duplex also can be affected by N6-methylation at the adenine residues. It has been reported that, at 15 °C, the methylated A·T base pair (mA·T) in the HMe-GATC duplex has a longer lifetime than does the unmethylated A·T base pair in the UMe-GATC duplex, indicating that the mA·T base pair is the less dynamic of the two. This characteristic dynamic feature of the HMe-GATC duplex, however, cannot solely explain the specific binding of SeqA to hemimethylated GATC site. In this study, which was performed under more physiological conditions (35 °C), we found that the two N6-methylations at the GATC site in the FMe-GATC duplex stabilized the two A·T base pairs (mA6·T19 and T7·mA18), such that their base-pair dissociation constants were half of those for the comparable base pairs in the hemimethylated GATC site and one-quarter of those for the comparable base pairs in the unmethylated site. On the other hand, the A·T base pairs in all GATC duplexes displayed similar opening rate constants (Table 2). This result indicates that the specific recognition by SeqA of the hemimethylated GATC site does not correlate with the dynamics of A·T base-pair opening in the GATC site.

However, we did observe unusual and provocative kinetic and dynamic properties associated with the two G·C base pairs (G5·C20 and C8·G17) that are adjacent to the A·T base pairs in the GATC site. αK_{op} values of these G·C base pairs revealed that their thermodynamic stabilities barely changed upon methylation of the central GATC site. However, the opening rate constants ($k_{op} = 1/\tau_0$) of the G5·C20 and C8·G17 base pairs in the HMe-GATC duplex were 3- to 8-fold larger than those of the corresponding base pairs in the FMe-GATC duplex (Table 2). Also, the base-pair closing rate constants ($k_{cl}/\alpha = k_{op}/\alpha K_{op}$) of the HMe-GATC duplex were 3- to 6-fold larger than those of the FMe-GATC duplex. These data indicate that the presence of two N6-methyl groups within the central GATC site dramatically reduces both the base-pair opening and closing rates of the neighboring G·C base pairs, compared to those of G·C base pairs adjacent to only a single N6-methylation. Accordingly, we suggest that the flexibility of the two G·C base pairs in the central GATC site plays a key role in the discrimination, by SeqA, between a hemimethylated and fully methylated GATC site, both of which contain methylated adenine residues. In contrast, discrimination between unmethylated and hemimethylated GATC sites by SeqA is based on the absence and presence of a methylated adenine residue.

The most striking dynamic feature we observed is related to the difference in base-pair dynamics of the C8·G17 base pair. Imino proton exchange occurs only when the opening angle of

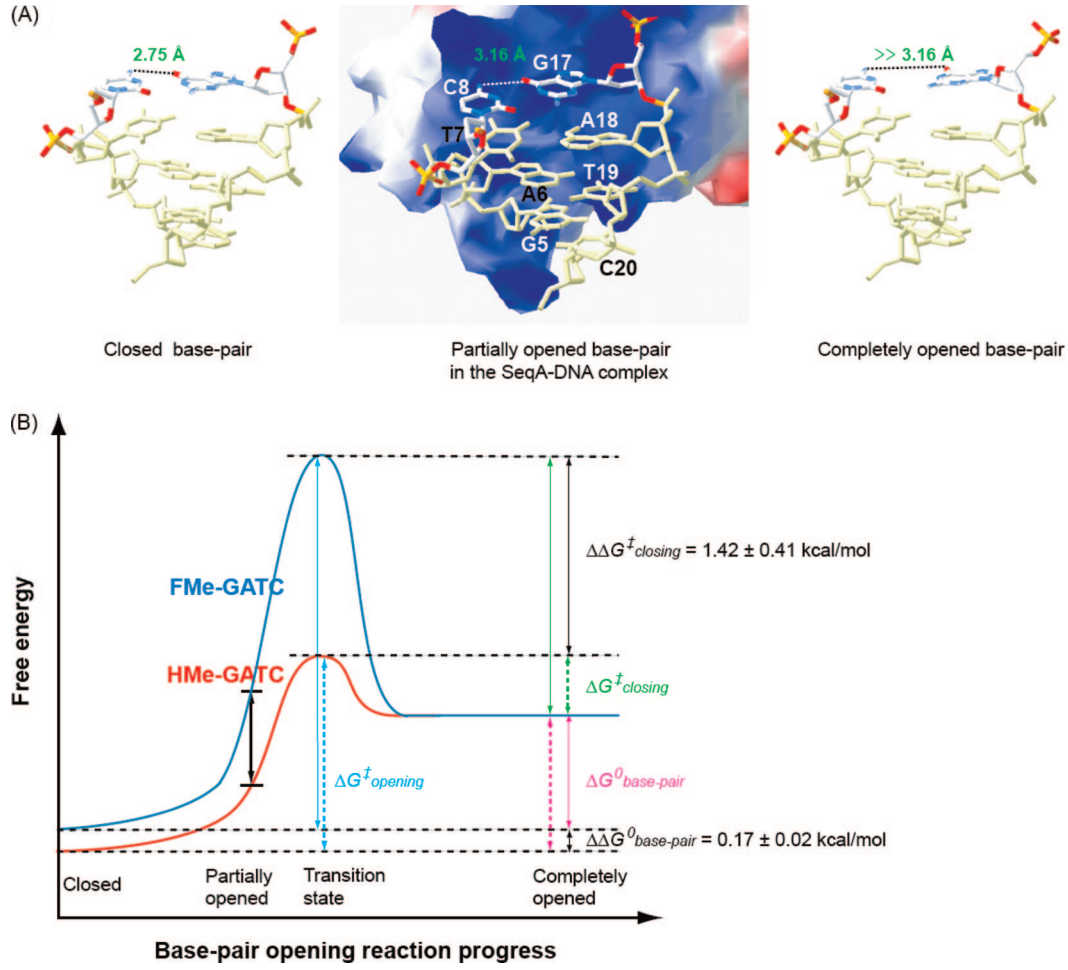

Figure 5. (A) Ball and stick views of the closed G·C base pair of the protein-free hemimethylated GATC duplex (left, PDB code 1UAB),[19] the partially opened G·C base pairs in the SeqA-hemimethylated GATC complex (middle, PDB code 1LRR)[16] and suggested completely opened G·C base pair (right), respectively. Hydrogen bonding heavy atom distances between guanine O6 and cytosine N4 are measured by a Swiss-pdb viewer.[44] (B) Schematic representations of the possible Gibbs free energy diagram of the base-pair opening and closing of the G·C base pair adjacent to the N6-methylated adenine residue in the HMe-GATC duplex (red) and the FMe-GATC duplex (blue). The magenta, cyan, and green lines indicate the $\Delta G^0_{\text{base pair}}$, $\Delta G^{\ddagger}_{\text{opening}}$, and $\Delta G^{\ddagger}_{\text{closing}}$, respectively. The dotted line represents parameters of HMe-GATC duplex, and the solid line represents those of FMe-GATC duplex.

the base pair is larger than 50°.[43] Figure 5A represents the three stages of the base-pair-opening reaction (closed base pair, partially opened base pair, and completely opened base pair). The Gibbs free energy difference between the closed and completely opened states of the C8·G17 base pair ($\Delta G^0_{\text{base pair}}$) can be calculated from the αK_{op} value using the equation $\Delta G^0_{\text{base pair}} = RT \ln(\alpha K_{\text{op}})$. The results revealed that the difference between the $\Delta G^0_{\text{base pair}}$ values of the C8·G17 base pair in HMe-GATC and FMe-GATC duplexes ($\Delta\Delta G^0_{\text{base pair}}$) is only 0.17 ± 0.02 kcal/mol. The $\Delta G^{\ddagger}_{\text{opening}}$ and $\Delta G^{\ddagger}_{\text{closing}}$ values (which represent energy differences between the closed and transition states, and the opened and transition states, respectively) can be evaluated from k_{op} and k_{cl} using the Arrhenius equation (see Table 2). Surprisingly, the $\Delta\Delta G^{\ddagger}_{\text{closing}}$ value was 1.42 ± 0.41 kcal/mol, more than 8 times larger than $\Delta\Delta G^0_{\text{base pair}}$. Figure 5B shows the possible Gibbs free energy diagrams for C8·G17 base-pair opening in the hemimethylated and fully methylated GATC duplexes, and Table 2 represents the Gibbs free energy values for each stage of the base-pair opening reaction and the equations used to calculate these values.

In the cocrystal structure of the hemimethylated GATC DNA duplex bound to the SeqA protein, the two G·C base pairs in the GATC site are partially opened, as evidenced by the observation that the hydrogen bonding heavy atom distance between G-O6 and C-N4 (3.1−3.3 Å) is longer than that of a Watson−Crick base pair (2.7−2.8 Å) (Figure 5A). This partially opened base-pair state is energetically less stable than a normal Watson−Crick base pair. Using free energy differences calculated from imino proton exchange experiments (Figure 5B, black solid arrow), we estimated the difference in energy that the G·C base pairs in the HMe- and FMe-GATC duplexes require to reach a partially opened state from a closed state. This energy difference was expected to be much larger than the difference between the $\Delta G^0_{\text{base pair}}$ values of the HMe- and FMe-GATC duplexes (∼0.17 ± 0.02 kcal/mol), because of the large energy difference between the transition states of both duplexes (∼1.42 ± 0.41 kcal/mol), which indicates that the FMe-GATC duplex needs more energy to form a partially opened base pair (this opening is thought to be essential for complex formation between the GATC site and the SeqA protein) (Figure 5B). In both the HMe-GATC and FMe-GATC duplexes, binding of SeqA protein to the DNA duplex can lower the total energy of the protein-DNA complex. However, the FMe-GATC duplex is expected to be significantly less stabilized by complex formation with SeqA, relative to the HMe-GATC duplex. Thus, we conclude that the hemimethylated GATC site is energetically

(43) Giudice, E.; Varnai, P.; Lavery, R. *ChemPhysChem* **2001**, *2*, 673–677.

more favorable for complex formation with SeqA than is the corresponding fully methylated complex.

Conclusion

We summarize the factors that determine the binding specificity of the SeqA protein to the hemimethylated GATC duplex in the background of numerous fully methylated sites as follows: (i) The unique groove structure of the hemimethylated GATC duplex disappears in the fully methylated one due to the additional N6-methylated adenine residue. (ii) The faster base-pair opening of the hemimethylated GATC duplex, relative to the fully methylated one, allows the hemimethylated GATC duplex to be recognized easily by the SeqA protein in the background of many fully methylated sites. And finally, (iii) the fact that less energy is required to form the partially opened state of the hemimethylated GATC site, relative to the fully methylated one, might explain why SeqA binds specifically to hemimethylated GATC sites.

Acknowledgment. We thank Dr. Hee-Chul Ahn for assistance with the NMR experiments. This study made use of the NMR facility at Korea Basic Science Institute, which is supported by Bio NMR Research Program of the Korean Ministry of Science and Technology (E28070). This work supported by a grant from the National Creative Research Initiative Program (to B.-S.C.) and a grant to EB-NCRC (R15-2003-012-01001-0) and the KOSEF Grant (R01-2007-000-10691-0) (to J.-H.L.) funded by the Korean Government (MEST). J.B. was supported by the BK21 project from Korean Government (MEST).

JA8038272

(44) Guex, N.; Peitsch, M. C. *Electrophoresis* **1997**, *18*, 2714–23.