

Systematic Investigations of Different Cytosine Modifications on CpG Dinucleotide Sequences: The Effects on the B-Z Transition

Shaoru Wang,[†] Yuelin Long,[†] Jiaqi Wang,[†] Yushu Ge,[†] Pu Guo,[†] Yi Liu,[†] Tian Tian,^{*,†} and Xiang Zhou^{*,†,‡}

[†]College of Chemistry and Molecular Sciences, Key Laboratory of Biomedical Polymers of Ministry of Education, Wuhan University, Wuhan, Hubei 430072, P. R. China

[‡]State Key Laboratory of Natural and Biomimetic Drugs, Peking University, Beijing 100191, P. R. China

Supporting Information

ABSTRACT: We have first demonstrated the distinctive effects of three newly reported epigenetic modifications, including 5hmC, 5fC, and 5caC, on B-Z transition of CpG dinucleotide DNAs. We have performed detailed assays and compared their effects. We further studied the regulation of B-Z transition of CpG dinucleotide dodecamers by alternating oxidation and alternating reduction.

left-handed DNA secondary structure,^{1,2} Z-DNA is generally stabilized in solutions with high concentrations of salts.³ Z-DNA has been considered to be a kinetically unstable conformation, with phosphate groups on the DNA backbones being more compact.⁴ Several important Z-DNA binding proteins have been discovered, and the biological roles of Z-DNA are also gradually being elucidated. 5-7 Thus far, it is widely accepted that negative supercoiling resulting from relocating RNA polymerase in transcription provides one potential for the stabilization of Z-DNA in vivo; Z-DNA's formation is closely linked with transcription level.⁸ It has also been reported that potential Z-DNA-forming sequences are highly dispersed near promoter regions of some mammalian genes.^{9,10} Repeated CpG dinucleotide sequences have been found to frequently occur in eukaryotic promoter DNA and adopt Z-DNA conformations more easily due to energetic favorability.¹¹ Due to the rigidity of Z-DNA, the formation of this structure was thought to interfere with its interaction with corresponding binding protein and/or transcription factors, thus leading to some subsequent biological effects.1

Cytosines in repeated CpG dinucleotides of vertebrate genomes could be methylated to form 5-methylcytosine (5mC in Figure 1) upon the functions of DNA methyltransferase (DNMT).^{13,14} CpG methylation plays an important role in



Figure 1. Structure illustration of cytosine and its derivatives including S-methylcytosine, S-hydroxymethylcysosine, S-formylcytosine, and S-carboxylcytosine.

epigenetics and is precisely regulated in different developmental stages and organs.^{15,16} Cytosine methylation has been found to facilitate the B-Z transition *in vitro*, and CpG methylation and the Z-DNA structure also influence each other *in vivo*.¹⁷ It was recently revealed that 5-hydroxymethylcytosine, 5-formylcytosine, and 5-carboxylcytosine (5hmC, 5fC, and 5caC, respectively, in Figure 1) could be generated through the oxidation of 5-methylcytosine by the ten-eleven translocation (TET) family of enzymes in mammals.^{18–21} These newly reported cytosine modifications, together with 5mC, are involved in multifaceted biological processes, such as embryonic development, the differentiation of embryonic stem cells, the regulation of genome functions, etc.²²

Until now, there have been no studies describing the B-Z transition of repeated CpG dinucleotide sequences containing these newly found cytosine modifications (including ShmC, 5fC, or 5caC) in mammalians. To investigate their important biological roles in genomic DNA, we asked whether the oxidized modifications of 5mC in CpG dinucleotide sequences would influence the B-Z transition process, including facilitating a transition like methylation or inhibiting the transition on opposite (Figure 2).

Because poly(dG-dC)-poly(dG-dC) (polyCG) polymers are commercially available and have been widely accepted as suitable model sequences to study the B-Z transition *in vitro*, we first attempted to investigate the B-Z transition behaviors of these derived polymers with homogenized modifications including



Figure 2. Varied oxidized modifications of SmC in CpG dinucleotide sequences: facilitating the B-Z transition similar to methylation or inhibiting the process.

Received: October 23, 2013 Published: December 23, 2013 hydroxymethylation, formylation, or carboxylation on all of the cytosines. Wells and Felsenfeld et al. reported the preparation of poly(dG-5mdC)-poly(dG-5mdC) (poly5mCG) in an efficient manner,^{23,24} which was also adopted by us. We incorporated different modified nucleoside triphosphates including 5-hydroxymethyl-2'-deoxycytidine-5'-triphosphate, 5-formyl-2'-deoxycytidine-5'-triphosphate, or 5-carboxyl-2'-deoxycytidine-5'-triphosphate into the synthetic polymers. After obtaining the purified products of synthetic poly(dG-5hmdC)-poly(dG-ShmdC) (polyShmCG), poly(dG-SfdC)-poly(dG-SfdC) (poly5fCG), poly(dG-5cadC)-poly(dG-5cadC) (poly5caCG), and poly5mCG, structural characterizations were performed. Thus, the synthetic polymers were digested into single nucleotides using nuclease S1, venom phosphodiesterase I (Type VI), and alkaline phosphatase and subjected to MS analysis.²⁵ Consistent with our predictions, peaks corresponding to the 5-hydroxymethyl-2'-deoxycytidine (Figure S1), 5-formyl-2'-deoxycytidine (Figure S2), or 5-carboxyl-2'-deoxycytidine (Figure S3) could be obtained, which strongly suggested the successful preparation of the desired polymers.

The typical Z-DNA was first observed and evidenced by circular dichroism (CD) spectroscopy²⁶ and was further characterized using X-ray crystallography after ~7 years.² To perform a comparison between the first prepared polymers (poly5hmCG, poly5fCG, and poly5caCG) and the already known polymers (polyCG and poly5mCG), the CD spectrum for each polymer was obtained at different concentrations of NaCl.^{27,28} For polyCG and poly5mCG, typical B-Z transitions could be identified as functions of increasing concentrations of NaCl using CD spectroscopy (Figure S4). In the case of poly5hmCG, CD spectra showed a dominant negative peak at \sim 250 nm and a positive peak at \sim 294 nm at 500 mM NaCl, indicating a typical B-DNA conformation. Surprisingly, the negative band of poly5hmCG at ~250 nm barely changed, even at 4.0 M NaCl, indicating its persistence in the B-DNA conformation in a high NaCl concentration, and no obvious Z-DNA could be characterized in CD (results in Figure 3). However, the CD spectrum of poly5fCG demonstrated a dominant positive peak at ~250 nm, a positive peak at ~270 nm and an evident negative peak at ~290 nm at 0 mM NaCl,



Figure 3. CD spectra of poly5hmCG, poly5fCG, and poly5caCG in 5 mM Na cacodylate buffer (pH 7.0 at 10 $^{\circ}$ C), at varied concentrations of NaCl ranging from 0 mM (control) to 4.0 M. All the polymers except poly5caCG (0.1 mM base concentration) were used at 0.4 mM base concentration.

indicating a typical Z-DNA conformation in very low-salt conditions (Figure 3). Based on our results, Z-DNA was more easily formed in formylated polymers compared with methylated polymers. Furthermore, the shape of the CD spectrum of poly5fCG could be identified as a mirror reflection of the shape of the CD spectrum of poly5hmCG, indicating a reverse conformational change from poly5hmCG to poly5fCG (results in Figure S5). The CD spectrum of poly5caCG also showed some characteristics of Z-DNA, which was further evidenced by a chiroptical probe of a known zinc(II) porphyrin 1 (Figures S6 and S7). We also calculated the midpoint NaCl concentrations for the B-Z transition of all the polymers (Table S2), which strongly suggested that the modifications of formylation and carboxylation at the 5' position facilitated the B-Z transition, while hydroxymethylation inhibited this process.

To further verify the hypothesis regarding the effects of cytosine modifications on B-Z transitions, we next investigated the B-Z transition behavior of chemically synthesized dodecamers, which contained only one modification site. 12mer-1hmC (X = 5hmC), 12mer-1fC (X = 5fC) and 12mer-1caC (X = 5caC)with the sequence of 5'- CG(X)GCGCGCG-3' were tested as the target sequences. Other DNA templates possessing the cytosine or 5mC existing at the corresponding X loci were used as controls (12mer-C, X = C; 12mer-1mC, X = 5mC). For 12mer-1mC, the stabilization of Z-DNA by incorporating 5mC has been observed; this finding was consistent with the above study. For 12mer-1hmC, a high concentration of NaCl (up to 2.0 M) would not produce an evident increase in the CD absorption at 250 nm. The B-Z transition of 12mer-1hmC could be observed only when NaCl was at 3.0 M, at which an obvious negative peak at \sim 294 nm (attributed to Z-DNA) appeared (results in Figure 4). In the



Figure 4. CD spectra of 12mer-C, 12mer-1mC, 12mer-1hmC, 12mer-1fC and 12mer-1caC in 5 mM Na cacodylate buffer (pH 7.0 at 10 °C), at varied concentrations of NaCl ranging from 0 mM (control) to 4.0 M (all oligomers, 10 μ M strand concentration).

case of 12mer-1fC, we observed a dominant negative peak at \sim 250 nm and a positive peak at \sim 294 nm in the CD spectra in 5.0 mM sodium cacodylate buffer, showing a B-DNA conformation. With addition of increasing amounts of NaCl, an obvious negative peak at \sim 294 nm was generated, indicating a typical B-Z transition. In our observations, the incorporation of 5fC possessing a formyl group into the dodecamer significantly stabilized the Z form: the CD spectrum of 12mer-1fC is approximately inverted in 2.0 M NaCl compared with the B-form

Journal of the American Chemical Society

spectrum observed in a low-salt solution. A similar B-Z transition was observed in 12mer-1caC. For the B-Z transition of dodecamers with only one modified cytosine, the order of the midpoint NaCl concentration was consistent with the above results related to the modified polymers, which further supported that natural cytosine modifications could exhibit different effects on the parent DNA (results in Table S2).

To further verify the effects of different modifications on DNA conformation, the dodecamers containing two modification sites including 5mC, 5hmC, 5fC, or 5caC were used in the subsequent assay. 12mer-2hmC ($\mathbf{X} = 5$ hmC, $\mathbf{Y} = 5$ hmC), 12mer-2fC ($\mathbf{X} = 5$ fC, $\mathbf{Y} = 5$ fC) and 12mer-2caC ($\mathbf{X} = 5$ caC, $\mathbf{Y} = 5$ caC) with the sequence of 5'- CG(\mathbf{X})GCGCG(\mathbf{Y})GCG-3' were tested as the target sequences. The 12mer-2mC possessing 5mC at the corresponding \mathbf{X} and \mathbf{Y} loci were used as a control (Figure S8). The conformation of 12mer-1hmC could be inverted at 3.0 M NaCl, while the B-Z transition of 12mer-2hmC was inhibited to an evident extent at the same concentration of NaCl (results in Figure 5). For 12mer-2hmC, a significantly higher B-Z transition



Figure 5. CD spectra of 12mer-2mC, 12mer-2hmC, 12mer-2fC and 12mer-2caC in 5 mM Na cacodylate buffer (pH 7.0 at 10 °C), at varied concentrations of NaCl ranging from 0 mM (control) to 4.0 M (all oligomers, 10 μ M strand concentration).

midpoint than that for 12mer-1hmC could be characterized in the assay (Table S2), which indicated a more significant B-DNA stabilization via more 5hmC modifications in the DNA. In the case of 12mer-2fC, the midpoint of the B-Z transition could be characterized at 1.5 M NaCl, which was lower than that of 12mer-2mC (results in Table S2). The 12mer-2fC with two 5fC modifications could be induced to a left-handed conformation with a better efficiency than 12mer-1fC with only one 5fC modification. Higher stabilizing effects could also be observed in 12mer-2caC compared with 12mer-1caC (table S2). Consistent with the results observed for poly5fCG and poly5caCG, double dominant positive peaks at \sim 250 and 270 nm appear in the CD spectra of 12mer-2fC and 12mer-2caC at high concentrations of NaCl. We could conclude that two modifications on cytosines could exert a more significant influence than one modification on their respective B-Z transitions in the short dodecamers. Additionally, the addition of a 5'-G on the end of the oligomers could inhibit the B-Z transition of the duplexes (results in Figure S9).

To gain more insights into the explicit effects of different modifications on cytosine, we analyzed the dodecamers with three or four modified cytosines scattered in the DNA sequence and compared them to those containing one or two modifications. 12mer-3mC (X = 5mC, Y = 5mC, Z = 5mC, Q= C), 12mer-4mC (X = 5mC, Y = 5mC, Z = 5mC, Q = 5mC), 12mer-3caC ($\mathbf{X} = 5caC$, $\mathbf{Y} = 5caC$, $\mathbf{Z} = 5caC$, $\mathbf{Q} = C$), and 12mer-4caC (X = 5caC, Y = 5caC, Z = 5caC, Q = 5caC) with the sequence of 5'-(X)G(Y)GCGCG(Z)G(Q)G-3' were tested as the target sequences. It was found that increased cytosine modifications including methylation and carboxylation further enhanced the B-Z transition (Figure S10). The CD spectra of 12mer-3mC and 12mer-4mC showed a similar shape at low concentrations of NaCl compared with 12mer-C. There was a significant positive peak at ~230 nm in the CD spectra of 12mer-3caC and 12mer-4caC, even in the absence of NaCl (Figure S10). Moreover, the CD absorbance of 12mer-3caC and 12mer-4caC at 294 nm decreased sharply upon increasing the concentration of NaCl, which was an important characteristic of Z-DNA formation.

The modification on cytosines could be accomplished using certain proteins under physiological conditions.²⁹ To further confirm the above effects of different cytosine modifications on the B-Z transition, *in situ* conversion and the following characterization using CD spectroscopy are necessary. It has been reported that the chemical conversion of 5fC to 5hmC could be achieved with a quantitative yield using a NaBH₄-based reduction in aqueous solution³⁰ (Figure 6A). Thus, we used



Figure 6. Reversible conversion between 12mer-2hmC and 12mer-2fC could be accomplished by NaBH₄-based alternate reduction or KRuO₄-based alternate oxidation. (A) Illustration of the oxidation—reduction cycle of DNAs; (B) the typical CD spectra of 12mer-2fC ($30 \ \mu g$ in $300 \ \mu L$) and its reduced products; (C) the typical CD spectra of 12mer-2hmC ($30 \ \mu g$ in $300 \ \mu L$) and its oxidized products.

12mer-2fC as a starting material and measured its CD spectrum, which showed a typical Z-DNA conformation. After the NaBH₄ reduction reaction, the purified DNA products were dissolved in 3.0 M NaCl and measured again using CD spectroscopy, which demonstrated a typical B-DNA conformation (Figure 6B). Thus, the conformation transition could be chemically regulated via a reductive reaction. Because the specific oxidation of 5hmC to 5fC could be achieved with potassium perruthenate (KRuO₄)³¹ (Figure 6A), 12mer-2hmC was oxidized and purified, and the subsequent CD study showed a typical Z-DNA conformation, consistent with our predictions (Figure 6C). A further reversible B-Z transition under high-salt concentrations could be repeatedly regulated from 12mer-2hmC or 12mer-2fC via chemical reactions up to five times (Figure S11).

Journal of the American Chemical Society

Many modifications on bases have been found to have effects on facilitating the B-Z transition.^{27,32} Here, we first demonstrated that DNA hydroxymethylation would inhibit the B-Z transition of CpG dinucleotide sequences and that this inhibition was significantly different from the effects of methylation. We found that 5fC and 5caC, the other potential oxidation intermediates of 5mC or 5hmC, could facilitate the B-Z transition. It has been revealed that 5fC or 5caC contained an intramolecular hydrogen bond,³³ thus shifting the tautomeric equilibrium toward a higher ratio of imino form (Figures S12 and S13).³⁴ The imino tautomers exhibiting only two hydrogen bonds were less stable than the amino ones with three hydrogen bonds, thus enhancing the B-Z transition of alternating CG duplexes. Since 5hmC did not contain 5-position carbonyl oxygen, there is no such weakened hydrogen bonding pattern,³⁵ as in 5fC or 5caC. This may explain the difference in B-Z propensity for the newly reported 5-C epigenetic modifications. Additionally, energetic and dynamics simulations have been performed via using biopolymer module of SYBYL 8.1 (Tripos, Inc.),³⁶ which were consistent with the experimental results (results in Table S3, Figures S14-S17). Because Z-DNA has been associated with gene activation, 7,37 we suggested that the oxidation states of 5mC may exert effects on gene expression via the perturbation of B-Z transitions. These new findings may provide new insights on DNA epigenetics and could represent a promising strategy in future epigenetic analysis.

ASSOCIATED CONTENT

Supporting Information

General methods and data. This information is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Authors

xzhou@whu.edu.cn ttian@whu.edu.cn

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

The authors thank the National Basic Research Program of China (973 Program) (2012CB720600, 2012CB720603), the National Science of Foundation of China (nos. 91213302, 21102108, 21372182, 81373256), the National Grand Program on Key Infectious Disease (2012ZX10003002-014), open funding of the State Key Laboratory of Bioorganic and Natural Products Chemistry, Shanghai Institute of Organic Chemistry, the Chinese Academy of Sciences, supported by Program for Changjiang Scholars and Innovative Research Team in University (IRT1030).

REFERENCES

(1) Mitsui, Y.; Langridge, R.; Shortle, B. E.; Cantor, C. R.; Grant, R. C.; Kodama, M.; Wells, R. D. *Nature* **1970**, 228, 1166.

(2) Wang, A. H.; Quigley, G. J.; Kolpak, F. J.; Crawford, J. L.; van Boom, J. H.; van der Marel, G.; Rich, A. *Nature* **1979**, *282*, 680.

(3) Drew, H.; Takano, T.; Tanaka, S.; Itakura, K.; Dickerson, R. E. *Nature* **1980**, *286*, 567.

(4) Callahan, L.; Han, F. S.; Watt, W.; Duchamp, D.; Kezdy, F. J.; Agarwal, K. Proc. Natl. Acad. Sci. U.S.A. **1986**, 83, 1617.

(5) Herbert, A.; Alfken, J.; Kim, Y. G.; Mian, I. S.; Nishikura, K.; Rich, A. Proc. Natl. Acad. Sci. U.S.A. **1997**, *94*, 8421.

(6) Rich, A.; Zhang, S. Nat. Rev. Genet. 2003, 4, 566.

(7) Kim, Y. G.; Muralinath, M.; Brandt, T.; Pearcy, M.; Hauns, K.; Lowenhaupt, K.; Jacobs, B. L.; Rich, A. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, 100, 6974.

(8) Champ, P. C.; Maurice, S.; Vargason, J. M.; Camp, T.; Ho, P. S. *Nucleic Acids Res.* **2004**, *32*, 6501.

(9) Thomas, M. J.; Freeland, T. M.; Strobl, J. S. *Mol. Cell. Biol.* **1990**, *10*, 5378.

(10) Maruyama, A.; Mimura, J.; Harada, N.; Itoh, K. *Nucleic Acids Res.* **2013**, *41*, 5223.

(11) Temiz, N. A.; Donohue, D. E.; Bacolla, A.; Luke, B. T.; Collins, J. R. *PLoS One* **2012**, *7*, e35558.

(12) Kim, Y. G.; Lowenhaupt, K.; Oh, D. B.; Kim, K. K.; Rich, A. Proc. Natl. Acad. Sci. U.S.A. 2004, 101, 1514.

(13) Dodge, J. E.; Ramsahoye, B. H.; Wo, Z. G.; Okano, M.; Li, E. *Gene* 2002, 289, 41.

(14) Okano, M.; Bell, D. W.; Haber, D. A.; Li, E. Cell 1999, 99, 247.

(15) Feng, J.; Zhou, Y.; Campbell, S. L.; Le, T.; Li, E.; Sweatt, J. D.; Silva, A. J.; Fan, G. Nat. Neurosci. 2010, 13, 423.

(16) Klysik, J.; Stirdivant, S. M.; Singleton, C. K.; Zacharias, W.; Wells, R. D. J. Mol. Biol. 1983, 168, 51.

(17) Evison, B. J.; Bilardi, R. A.; Chiu, F. C.; Pezzoni, G.; Phillips, D. R.; Cutts, S. M. Nucleic Acids Res. **2009**, *37*, 6355.

(18) Tahiliani, M.; Koh, K. P.; Shen, Y.; Pastor, W. A.; Bandukwala, H.; Brudno, Y.; Agarwal, S.; Iyer, L. M.; Liu, D. R.; Aravind, L.; Rao, A. *Science* **2009**, *324*, 930.

(19) Ito, S.; D'Alessio, A. C.; Taranova, O. V.; Hong, K.; Sowers, L. C.; Zhang, Y. *Nature* **2010**, *466*, 1129.

(20) Ito, S.; Shen, L.; Dai, Q.; Wu, S. C.; Collins, L. B.; Swenberg, J. A.; He, C.; Zhang, Y. *Science* **2011**, 333, 1300.

(21) He, Y. F.; Li, B. Z.; Li, Z.; Liu, P.; Wang, Y.; Tang, Q.; Ding, J.; Jia, Y.; Chen, Z.; Li, L.; Sun, Y.; Li, X.; Dai, Q.; Song, C. X.; Zhang, K.; He, C.; Xu, G. L. *Science* **2011**, 333, 1303.

(22) Pfaffeneder, T.; Hackner, B.; Truss, M.; Munzel, M.; Muller, M.; Deiml, C. A.; Hagemeier, C.; Carell, T. *Angew. Chem., Int. Ed. Engl.* **2011**, *50*, 7008.

(23) Wells, R. D.; Larson, J. E.; Grant, R. C.; Shortle, B. E.; Cantor, C. R. J. Mol. Biol. 1970, 54, 465.

(24) Behe, M.; Felsenfeld, G. Proc. Natl. Acad. Sci. U.S.A. 1981, 78, 1619.

(25) Song, C. X.; Szulwach, K. E.; Fu, Y.; Dai, Q.; Yi, C.; Li, X.; Li, Y.; Chen, C. H.; Zhang, W.; Jian, X.; Wang, J.; Zhang, L.; Looney, T. J.; Zhang, B.; Godley, L. A.; Hicks, L. M.; Lahn, B. T.; Jin, P.; He, C. *Nat. Biotechnol.* **2011**, *29*, 68.

(26) Pohl, F. M.; Jovin, T. M. J. Mol. Biol. 1972, 67, 375.

(27) Xu, Y.; Ikeda, R.; Sugiyama, H. J. Am. Chem. Soc. 2003, 125, 13519.

(28) Kypr, J.; Kejnovska, I.; Renciuk, D.; Vorlickova, M. Nucleic Acids Res. 2009, 37, 1713.

(29) Franchini, D. M.; Schmitz, K. M.; Petersen-Mahrt, S. K. Annu. Rev. Genet. 2012, 46, 419.

(30) Song, C. X.; Szulwach, K. E.; Dai, Q.; Fu, Y.; Mao, S. Q.; Lin, L.; Street, C.; Li, Y.; Poidevin, M.; Wu, H.; Gao, J.; Liu, P.; Li, L.; Xu, G. L.; Jin, P.; He, C. *Cell* **2013**, *153*, 678.

(31) Booth, M. J.; Branco, M. R.; Ficz, G.; Oxley, D.; Krueger, F.; Reik, W.; Balasubramanian, S. Science **2012**, 336, 934.

(32) Sugiyama, H.; Kawai, K.; Matsunaga, A.; Fujimoto, K.; Saito, I.; Robinson, H.; Wang, A. H. *Nucleic Acids Res.* **1996**, *24*, 1272.

(33) Munzel, M.; Lischke, U.; Stathis, D.; Pfaffeneder, T.; Gnerlich, F. A.; Deiml, C. A.; Koch, S. C.; Karaghiosoff, K.; Carell, T. *Chemistry* **2011**, *17*, 13782.

(34) Hashimoto, H.; Hong, S.; Bhagwat, A. S.; Zhang, X.; Cheng, X. *Nucleic Acids Res.* **2012**, *40*, 10203.

(35) Hu, J.; Chen, Y.; Xu, X.; Wu, F.; Xing, X.; Xu, Z.; Xu, J.; Weng, X.; Zhou, X. *Bioorg. Med. Chem. Lett.* **2013**, DOI: 10.1016/j.bmcl.2013.11.017

(36) Leonard, G. A.; McAuley-Hecht, K.; Brown, T.; Hunter, W. N. Acta Crystallogr., Sect. D: Biol. Crystallogr. 1995, 51, 136.

(37) Oh, D. B.; Kim, Y. G.; Rich, A. Proc. Natl. Acad. Sci. U.S.A. 2002, 99, 16666.