Effect of epigenetic modifications on the formation of non-canonical nucleic acid structures

Wataru Yoshida, School of Bioscience and Biotechnology, Tokyo University of Technology
Taiji Oyama, Sales Division, JASCO Corporation

Abstract

Non-canonical nucleic acid structures such as triplexes, stem loops, D-loops, R-loops, cruciforms, G-quadruplexes (G4), and intercalated motifs (i-motifs) are formed in genomic DNA and RNA. In cells, non-canonical nucleic acid structures regulate biological processes; however, little is known regarding the mechanism via which the formation of non-canonical nucleic acid structures is regulated. In mammalian cells, the most common DNA and RNA modifications are 5-methylcytosine in CpG dinucleotides and N^6-methyladenine, respectively. These modifications enhance stacking interactions with neighboring bases because the methyl group increases the molecular polarizability of the bases. In addition, 5-methylcytosine increases the base pairing energy of C-C^+ formed in i-motif structures, and N^6-methyladenine inhibits A-T base pairing, suggesting that these modifications affect the formation and thermal stability of noncanonical nucleic acid structures. Here, we describe the effect of 5-methylcytosine modification on *VEGF* G4 and i-motif structures, and that of N^6-methyladenine modification on the G4 structure formed by GGA repeat RNA.

Keywords: G-quadruplex, intercalated motif, 5-methylcytosine, N^6-methyladenine, circular dichroism spectrum analysis

1. Introduction

The human body is composed of approximately 37 trillion cells of 200 different types. These diverse cells develop from a single fertilized egg cell. During the development of the fertilized egg cell to each tissue, genomic DNA is replicated accurately. Therefore, all cells from the same individual harbor identical genomic DNA; however, the functions of each cell differ. Why do cells containing the same genome exhibit different functions? This is because the gene expression patterns differ with each cell type. Human genomic DNA contains approximately 20,000 genes, and epigenetic modification is one of the factors that control tissue-specific gene expression. Epigenetics is defined as "the study of changes in gene function that are heritable and that do not entail a change in DNA sequence" [1]. DNA methylation is an important epigenetic modification.

In mammalian cells, DNA methylation mainly occurs at the fifth position of cytosine in the CpG dinucleotide sequence (Fig. 1A)[2]. Human genomic DNA contains approximately 28 million CpG sequences[3]. Methyl-CpG binding domain proteins bind to methyl-CpG and recruit chromatin remodeling factors to induce heterochromatin formation. Therefore, gene expression is generally suppressed when the CpGs in their promoters are methyl-

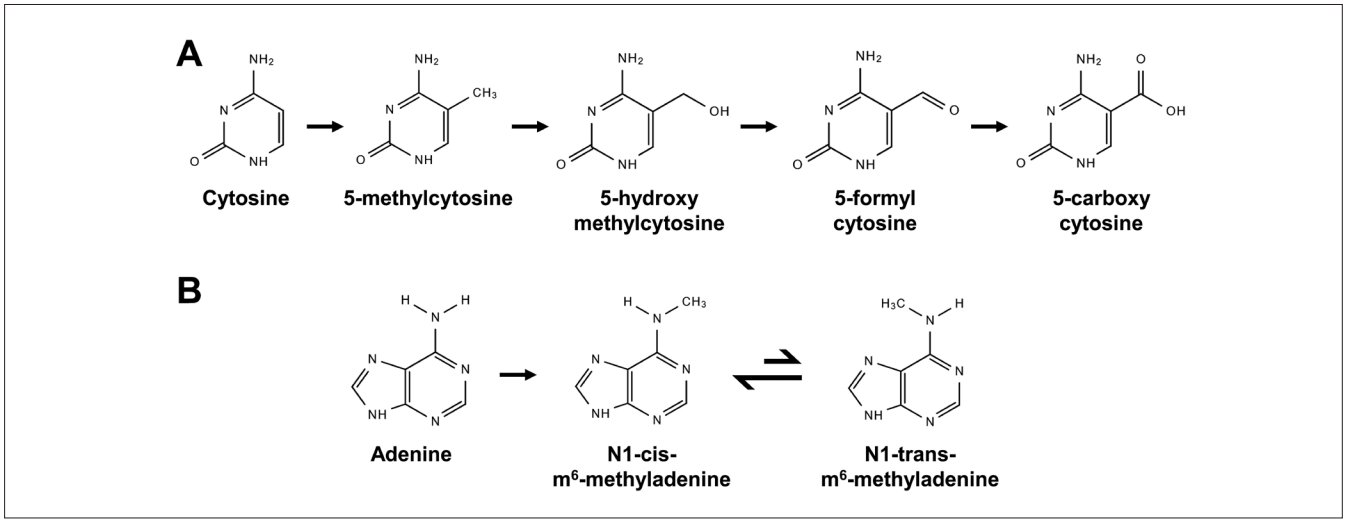

Fig. 1 Methylation of cytosine on CpG and adenine. (A) The fifth position of cytosine in CpG is methylated to form 5-methylcytosine. During the TET-mediated active DNA demethylation process, 5-methylcytosine is sequentially oxidized to 5-hydroxymethylcytosine, 5-formylcytosine, and 5-carboxylcytosine. (B) The sixth position of adenine is methylated to form N^6-methyladenine. N^6-methyladenine has two rotational isomers, *cis* and *trans*, formed in a 19:1 ratio.

ated. Thus, CpG methylation functions as a switch that controls gene expression. Tissue-specific methylation patterns are established during development. When induced pluripotent stem cells are produced from somatic cells, a methylation pattern similar to that of the embryonic stem cells is formed, indicating that the CpG methylation pattern can be used as a marker to determine cellular characteristics[4].

In addition to 5mC, other modified bases such as 5-hydroxymethylcytosine, 5-formylcytosine, and 5-carboxylcytosine, the intermediate products of the ten-eleven translocation (TET)-mediated active DNA demethylation process, have been identified in mammalian cells (Fig. 1A)[5, 6]. RNA also undergoes various modifications, with the most abundant modified base being N^6-methyladenine (m^6A), in which the sixth position of adenine is methylated (Fig. 1B)[7]. m^6A has two rotational isomers, *cis* and *trans*, where the methyl group is located on the N^1 and N^7 sides, respectively[8, 9]. The *cis* and *trans* isomers are formed in the ratio of 19:1, with the *cis* form being formed predominantly. In the *cis* form, the methyl group inhibits the formation of Watson-Crick base pairs with thymine. Studies have shown that the secondary structure of mRNA is changed by m^6A modification to control splicing[10], indicating that regulation of the formation of non-canonical nucleic acid structures by epigenetic modifications regulates biological functions.

DNA usually forms a double helical structure; however, it also forms non-canonical structures, such as stem loops, triplexes, G-quadruplexes (G4), and intercalated motifs (i-motifs)[11]. G4 is a four-stranded nucleic acid structure formed by the stacking of the G:G:G:G tetrad. The formation of G4 structures requires ions, such as K^+ coordinated between the G:G:G:G tetrad and Na^+ coordinated within the G:G:G:G tetrad. G4 structures were classified into three types based on the strand orientation: parallel, antiparallel, and hybrid (Fig. 2A-C). Additionally, ribonucleotides with four repeated GGA sequences ($r(GGA)_4$) formed a unique dimeric G4 structure containing the G:G:G:G tetrad and G(:A):G:G(:A):G hexad structures (Fig. 2D). The i-motif is a four-stranded nucleic acid structure formed by intercalated hemiprotonated cytosine base pairs ($C\text{-}C^+$) (Fig. 2E). As the formation of the i-motif structure requires protonation of cytosine, the i-motif usually forms under acidic conditions.

In genomic DNA, G4-forming sequences are found in guanine-rich regions, and their cytosine-rich complementary strands contain i-motif-forming sequences. In human genomic DNA, G4 structure-forming motifs ($G_3N_{1\text{-}7}G_3N_{1\text{-}7}G_3N_{1\text{-}7}G_3$) are frequently found near transcription start sites[12]. In fact, comprehensive analysis using DNA microarrays and next-generation sequencing with G4 ligands have revealed that G4-forming sequences are frequently found near transcription start sites[13–15]. G4 structure formation may be involved in transcriptional regulation; however, little is known about the mechanism of formation of G4 and i-motif structures from double helix structures.

We analyzed the effects of epigenetic modifications on the formation of non-canonical nucleic acid structures. Methyl groups are electron donors; therefore, methylation at the fifth position of cytosine enhances stacking interactions by increasing the molecular polarizability of pyrimidine[16]. Methylation also increases the base-pairing energy of the $C\text{-}C^+$ base pairs[17]. While m^6A enhances stacking interactions with neighboring bases, it inhibits Watson-Crick base pair formation as described above[18]. Therefore, the effects of epigenetic modifications on noncanonical nucleic acid structures can be elucidated by measuring the thermal stability of the modified noncanonical nucleic acid structures.

Dimethyl sulfate footprinting, circular dichroism (CD) spectrum analysis, nuclear magnetic resonance, and X-ray crystallography have been used to analyze non-canonical nucleic acid structures. Among these methods, CD spectral analysis is the most convenient. Since G4 and i-motif structures exhibit characteristic CD spectra, the thermal stability of these structures can be measured conveniently and precisely under various conditions. In this study, we describe the effects of epigenetic modifications on the thermal stability of the G4 and i-motif structures revealed using CD spectrum analysis.

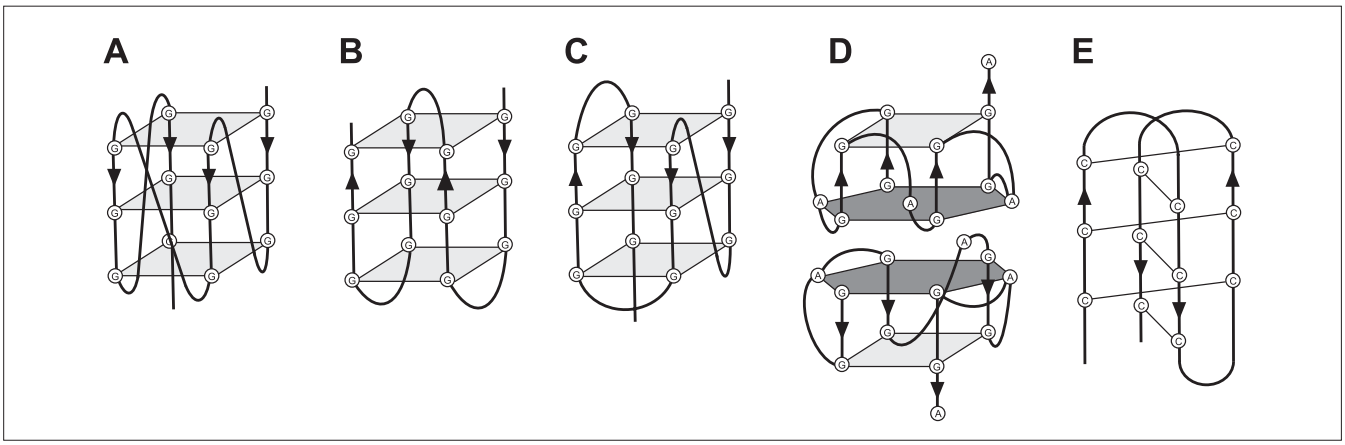

Fig. 2 Secondary structures of G4 and i-motif. (A) Parallel G4, (B) antiparallel G4, (C) hybrid G4, (D) dimeric G4 formed by $r(GGA)_4$, and (E) i-motif structure.

2. CD spectrum analysis for elucidating the effect of modified bases on non-canonical nucleic acid structures

We analyzed the effect of modified bases on the G4[19] and i-motif structures[20] present in the promoter of the vascular endothelial growth factor (*VEGF*) gene and the dimeric G4 structure formed by ribonucleotides with four repeated GGA sequences (r(GGA)$_4$)[21]. CpG-methylated *VEGF* G4, i-motif structure-forming deoxyribonucleotides, and m^{6}A-modified r(GGA)$_4$ were purchased from Tsukuba Oligo Service. The oligonucleotide sequences are listed in Table 1.

Table1 Sequences of oligonucleotides used in this study

	Sequence
VEGF G4	5′−CGGGGCGGGCCGGGGGCGGGG−3′
VEGF i-motif	5′−CCCCGCCCCCGGCCCGCCCCG−3
r(GGA)$_4$	5′−GGAGGAGGAGGA−3′

Methylated bases are underlined. 5mC was used for *VEGF* G4 and the i-motif. m^{6}A was used for r(GGA)$_4$.

2-1 Effect of CpG methylation on the formation of the *VEGF* G4 structure

The *VEGF* promoter contains a G4 structure-forming sequence and its complementary strand contains an i-motif structure-forming sequence (Table 1). We have previously shown that the polymerase chain reaction (PCR) amplification efficiency of the *VEGF* promoter region decreased when the CpG of this region was methylated[22]. The DNA polymerase is arrested at G4 structures formed on template DNA, suggesting that *VEGF* G4 and i-motif structures are stabilized by CpG methylation. As the *VEGF* G4 and i-motif structure-forming sequences contain four CpG sites, all CpG-methylated oligonucleotides were synthesized.

The PCR solution contained Na^+ and Mg^{2+} but not K^+, suggesting that CpG methylation increased the thermal stability of the *VEGF* G4 structure under these ionic conditions. The oligonucleotide solutions (20 μM), prepared in a buffer containing 5 mM NaH_2PO_4, 9.25 mM Na_2HPO_4, and 5 mM $MgCl_2$ (pH 7.4), were incubated at 95 ℃ for 3 min using a thermal cycler (Applied Biosystems Veriti Thermal Cycle, Thermo Fisher Scientific) and then gradually cooled from 95 ℃ to 25 ℃ over 30 min. The oligonucleotide solution (120 μL) was added to a cuvette of 1 mm optical path length (JASCO), followed by the addition of a mineral oil (40 μL, Bio-Rad). The cuvette was set in a J-1500 spectropolarimeter (JASCO) and the CD spectra were obtained from 320 nm to 220 nm as the temperature (cell holder temperature) was increased from 25 ℃ to 95 ℃ at 5 ℃ intervals.

Unmethylated and methylated *VEGF* G4 structure-forming oligonucleotides showed CD spectra with a positive peak at around 263 nm and a negative peak at around 245 nm at 25 ℃ (Fig. 3A, B), indicating that these oligonucleotides mainly form parallel G4 structures in the presence of Na^+ and Mg^{2+}. The positive peak at 263 nm decreases with increasing temperature, indicating that the thermal stability of the G4 structure can be evaluated using the CD value at 263 nm. Next, samples were prepared as described above and CD values at 263 nm were measured with increasing the temperature from 25 ℃ to 95 ℃ at 1 ℃ intervals. The CD values at 25 ℃ and 95 ℃ were set to 100 % and 0 %, respectively, and then used to normalize the data (Fig. 3C). The melting temperature (T_m) of each G4 structure was calculated by fitting the melting curves to a two-phase transition model using Python[23]. The T_m value of unmethylated and methylated *VEGF* G4 structures were calculated to be 69.2 ± 1.6 ℃ and 74.6 ± 1.1 ℃, respectively. These results demonstrated that CpG methylation increased the T_m value of *VEGF* G4 structure by 5.4 ℃.

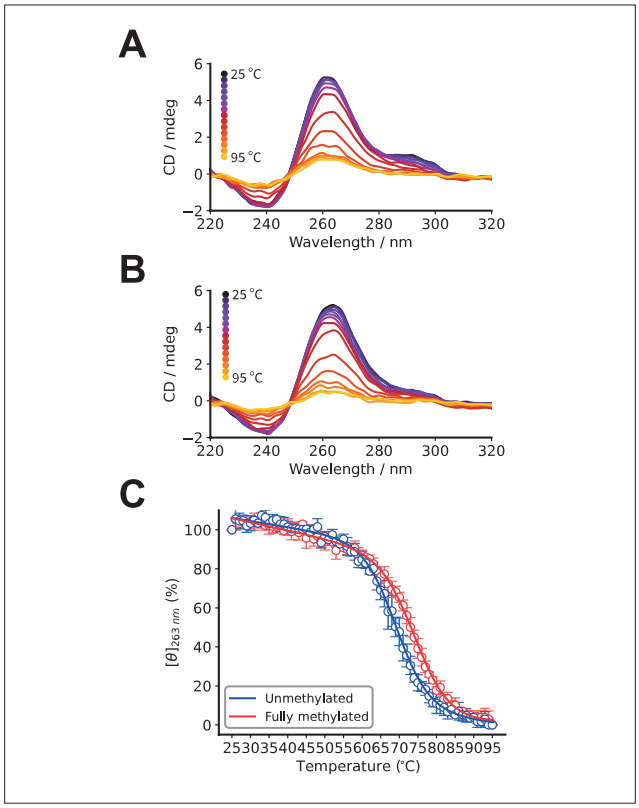

Fig. 3 Effect of CpG methylation on the thermal stability of the *VEGF* G4 structure. CD spectra of 20 μM unmethylated and methylated *VEGF* G4 structure-forming oligonucleotides were measured from 25℃ to 95℃ in a buffer containing 5 mM NaH_2PO_4, 9.25 mM Na_2HPO_4, and 5 mM $MgCl_2$ (pH 7.4). (A) CD spectra of unmethylated *VEGF* G4 structure-forming oligonucleotide, (B) CD spectra of methylated *VEGF* G4 structure-forming oligonucleotide, and (C) CD melting curves of unmethylated and methylated *VEGF* G4 structures. Reprinted from ref. 19. Copyright (2018), with permission from Elsevier.

2-2 Effect of CpG methylation on the formation of *VEGF* i-motif structure

As the formation of the i-motif structure requires protonation of cytosine, the i-motif usually forms under acidic conditions. Unmethylated and methylated *VEGF* i-motif

structure-forming deoxyribonucleotides were prepared at 20 μM in 25 mM sodium cacodylate (pH 5.0). Oligonucleotide solutions were incubated at 95 ℃ for 3 min and then gradually cooled from 95 ℃ to 25 ℃ over 30 min. The oligonucleotide solution (120 μL) was added to a cuvette of 1 mm optical path length, followed by the addition of a mineral oil (40 μL). The CD spectra were obtained from 320 nm to 220 nm as the temperature (cell holder temperature) was increased from 25 ℃ to 95 ℃ at 1 ℃ intervals using a J-1500 spectropolarimeter.

The unmethylated and methylated *VEGF* i-motif structure-forming oligonucleotides showed CD spectra with a positive peak at around 290 nm and a negative peak at around 265 nm at 25 ℃, which is the characteristic spectra of the i-motif structure (Fig. 4A, B). CD values at 290 nm at 25 ℃ and 95 ℃ were set to 100 % and 0 %, respectively, and then used to normalize the data to calculate the T_m value of each i-motif structure as described above (Fig. 4C). The T_m value of unmethylated and methylated *VEGF* i-motif structures were calculated to be 62.3 ± 0.2 ℃ and 65.7 ± 0.3 ℃, respectively. These results demonstrated that the T_m value of the *VEGF* i-motif structure increased by 3.4 ℃ due to CpG methylation.

2-3 Effect of m^{6}A modification on the formation of the G4 structure formed by r(GGA)$_4$

The *c-Myb* promoter contains GGA repeats that form dimeric G4 structures. Deoxyribonucleotide d(GGA)$_4$, which consists of four repeated GGA sequences, forms a dimeric G4 structure that includes G:G:G:G tetrads and G(:A):G(:A):G(:A):G heptads [24]. Ribonucleotide r(GGA)$_4$, which consists of four repeated GGA sequences, forms a dimeric G4 structure that includes G:G:G:G tetrads and G(:A):G:G(:A):G hexads [25]. In the G4 structure formed by r(GGA)$_4$, the hydrogen at the N^6 position of the third adenine (A3) was involved in hydrogen bond formation. Therefore, we hypothesized that the m^{6}A modification at A3 would decrease the thermal stability of the G4 structure. r(GGA)$_4$ with m^{6}A modification at A3 was prepared at a concentration of 20 μM in a buffer containing 4.8 mM NaH_2PO_4, 5.2 mM Na_2HPO_4, 10 mM NaCl, and 100 mM KCl (pH 6.7). The oligonucleotide solutions were incubated at 95 ℃ for 3 min and then gradually cooled from 95 ℃ to 25 ℃ over 30 min. The oligonucleotide solution (120 μL) was added to a cuvette of 1 mm optical path length, followed by the addition of a mineral oil (40 μL). The CD spectra were obtained from 320 nm to 220 nm as the temperature (cell holder temperature) was increased from 25 ℃

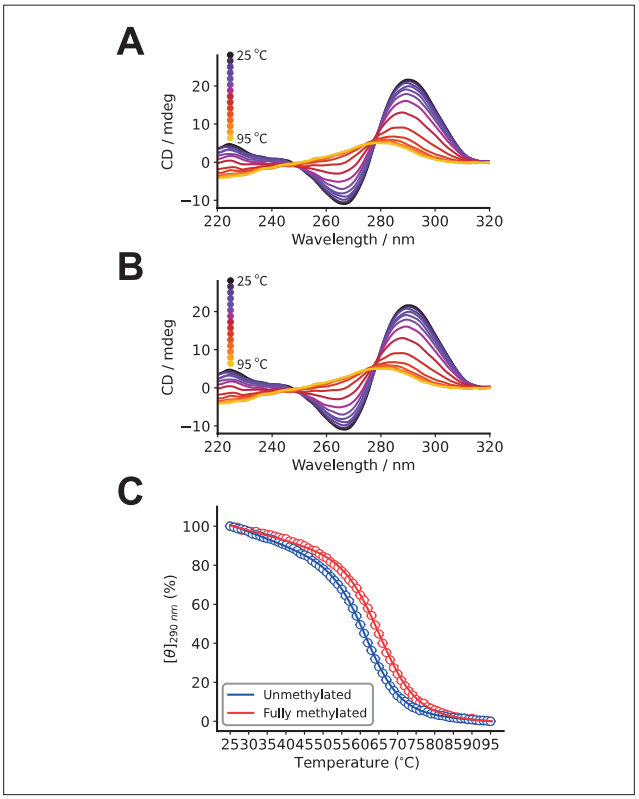

Fig. 4 Effect of CpG methylation on the thermal stability of the *VEGF* i-motif structure. The CD spectra of 20 μM unmethylated and methylated *VEGF* i-motif structure-forming oligonucleotides were measured from 25℃ to 95 ℃ in 25 mM sodium cacodylate (pH 5.0). (A) CD spectra of unmethylated *VEGF* i-motif structure-forming oligonucleotide, (B) CD spectra of methylated *VEGF* i-moitf structure-forming oligonucleotide, and (C) CD melting curves of unmethylated and methylated *VEGF* i-motif structures. Reprinted from ref. 20. Copyright (2022), with permission from Elsevier.

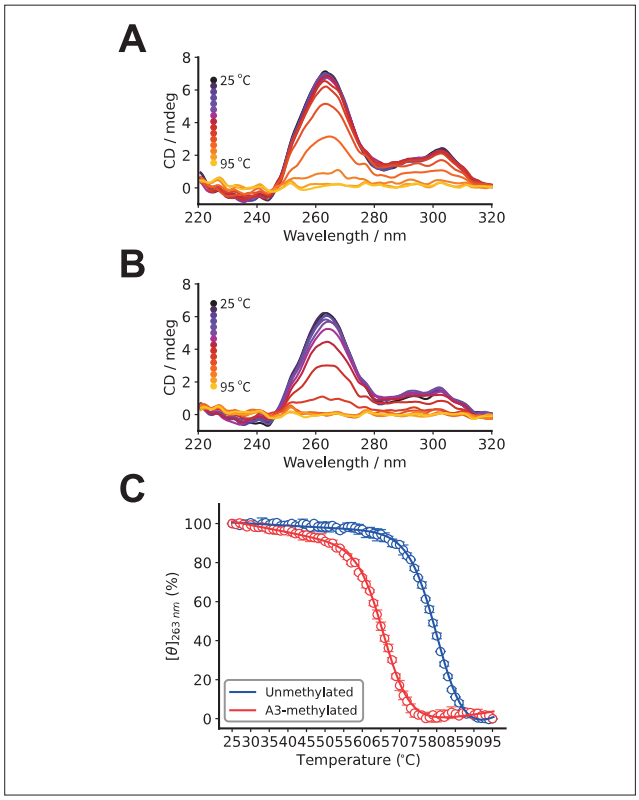

Fig. 5 Effect of m^{6}A modification on thermal stability of G4 structure formed by r(GGA)$_4$. The CD spectra of 20 μM unmethylated and m^{6}A-modified r(GGA)$_4$ were determined from 25 ℃ to 95 ℃ in a buffer containing 4.8 mM NaH_2PO_4, 5.2 mM Na_2HPO_4, 10 mM NaCl, and 100 mM KCl (pH 6.7). (A) CD spectra of unmethylated r(GGA)$_4$, (B) CD spectra of m^{6}A-modified r(GGA)$_4$, and (C) CD melting curves of unmethylated and m^{6}A-modified G4 structure formed by r(GGA)$_4$. Reprinted from ref. 21. Copyright (2022), with permission from Elsevier.

to 95 ℃ at 1 ℃ intervals using a J-1500 spectropolarimeter.

Unmethylated and m^{6}A-modified r(GGA)$_4$ showed a characteristic CD spectra for the dimeric G4 structure, with a positive peak at 263 nm at 25 ℃ (Fig. 5A, B). The CD values at 263 nm at 25 ℃ and 95 ℃ were set to 100 % and 0 %, respectively, and then used to normalize the data to calculate T_m value of each G4 structure formed by r(GGA)$_4$ (Fig. 5C). The T_m value of unmethylated and m^{6}A-modified G4 structures formed by r(GGA)$_4$ were calculated to be 80.8 ± 0.5 ℃ and 66.1 ± 0.4 ℃ , respectively. These results demonstrated that the T_m value of the G4 structure formed by r(GGA)$_4$ decreased by 14.7 ℃ due to the m^{6}A modification at A3.

3. Conclusions

In this study, we described the effect of epigenetic modifications on the thermal stability of *VEGF* G4, *VEGF* i-motif, and the G4 formed by r(GGA)$_4$. We also reported that CpG methylation affects the thermal stability of *c-kit*2 G4 structures [26] and that the m^{6}A modification affects the thermal stability of *c-kit*1 G4 structures [27], telomere G4 structures [28], and G4 structures formed by d(GGA)$_8$ and d(GGA)$_4$ [21]. Additionally, the effects of 5-hydroxymethylcytosine, 5-formylcytosine, and 5-carboxylcytosine on the formation of noncanonical structures were analyzed. CD spectrum analysis is not only easy to perform but also provides highly reproducible and accurate data; therefore, the effects of epigenetic modifications on various noncanonical structures can be analyzed. Genomic DNA and RNA contain more than 17 and 160 types of modified bases, respectively [29]. In the future, novel biological processes mediated by non-canonical structures, the formation of which is regulated by epigenetic modifications, will be elucidated, and the importance of CD spectrum analysis will continue to increase.

● Reference

1) C. Wu and J. R. Morris, *Science*, 2001, **293**, 1103–1105.
2) M. V. C. Greenberg and D. Bourc'his, *Nat. Rev. Mol. Cell Biol*., 2019, **20**, 590–607.
3) Z. D. Smith and A. Meissner, *Nat. Rev. Gent.*, 2013, **14**, 204–220.
4) R. Lister, M. Pelizzola, Y. S. Kida, R. D. Hawkins, J. R. Nery, G. Hon, J. Antosiewicz-Bourget, R. O'Malley, R. Castanon, S. Klugman, M. Downes, R. Yu, R. Stewart, B. Ren, J. A. Thomson, R. M. Evans and J. R. Ecker, *Nature*, 2011, **471**, 68–73.
5) M. Tahiliani, K. P. Koh, Y. Shen, W. A. Pastor, H. Bandukwala, Y. Brudno, S. Agarwal, L. M. Iyer, D. R. Liu, L. Aravind and A. Rao, *Science*, 2009, **324**, 930–935.
6) S. Ito, L. Shen, Q. Dai, S. C. Wu, L. B. Collins, J. A. Swenberg, C. He and Y. Zhang, *Science*, 2011, **333**, 1300–1303.
7) F. Qu, P. S. Tsegay and Y. Liu, *Front. Mol. Biosci.*, 2021, **8**, 645823.
8) J. D. Engel and P. H. von Hippel, *Biochemistry*, 1974, **13**, 4143–4158.
9) J. D. Engel and P. H. von Hippel, *J. Biol. Chem*., 1978, **253**, 927–934.
10) N. Liu, Q. Dai, G. Zheng, C. He, M. Parisien and T. Pan, *Nature*, 2015, **518**, 560–564.
11) R. C. Duardo, F. Guerra, S. Pepe and G. Capranico, *Biochimie*, 2023, **214**, 176–192.
12) J. L. Huppert and S. Balasubramanian, *Nucleic Acids Res*., 2005, **33**, 2908–2916.
13) **K. Iida, T. Nakamura, W. Yoshida, M. Tera, K. Nakabayashi**, K. Hata, K. Ikebukuro and K. Nagasawa, *Angew. Chem. Int. Ed. Engl*., 2013, **52**, 12052–12055.
14) V. S. Chambers, G. Marsico, J. M. Boutell, M. Di Antonio, G. P. Smith and S. Balasubramanian, *Nat. Biotechnol*., 2015, **33**, 877–881.
15) W. Yoshida, H. Saikyo, K. Nakabayashi, H. Yoshioka, D. H. Bay, K. Iida, T. Kawai, K. Hata, K. Ikebukuro, K. Nagasawa and I. Karube, *Sci. Rep*., 2018, **8**, 3116.
16) L. C. Sowers, B. R. Shaw and W. D. Sedwick, *Biochem. Biophys. Res. Commun*., 1987, **148**, 790–794.
17) B. Yang and M. T. Rodgers, *J. Am. Chem*. Soc., 2014, **136**, 282–290.
18) C. Roost, S. R. Lynch, P. J. Batista, K. Qu, H. Y. Chang and E. T. Kool, *J. Am. Chem. Soc*., 2015, **137**, 2107–2115.
19) W. Yoshida, M. Terasaka, S. Laddachote and I. Karube, *Biochim. Biophys. Acta Gen. Subj.*, 2018, **1862**, 1933–1937.
20) K. Kimura, D. Oshikawa, K. Ikebukuro and W. Yoshida, *Biochem. Biophys. Res. Commun*., 2022, **594**, 88–92.
21) Y. Iwasaki, Y. Ookuro, K. Iida, K. Nagasawa and W. Yoshida, *Biochem. Biophys. Res. Commun*., 2022, **597**, 134–139.
22) **W. Yoshida, H. Yoshioka, D. H. Bay, K. Iida, K. Ikebukuro**, K. Nagasawa and I. Karube, *Anal. Chem*., 2016, **88**, 7101–7107.
23) M. Moriya, T. Oyama, M. Goto, K. Ikebukuro, W. Yoshida, *STAR Protoc*., 2025, **6**, 103646.
24) A. Matsugami, K. Ouhashi, M. Kanagawa, H. Liu, S. Kanagawa, S. Uesugi and M. Katahira, *J. Mol. Biol*., 2001, **313**, 255–269.
25) **T. Mashima, A. Matsugami, F. Nishikawa, S. Nishi**kawa and M. Katahira, *Nucleic Acids Res*., 2009, **37**, 6249–6258.
26) S. Laddachote, R. Ishii and W. Yoshida, *BBA Advances*, 2021, **1**, 100007.
27) S. Laddachote, M. Nagata and W. Yoshida, *Biochem. Biophys. Res. Commun*., 2020, **524**, 472–476.
28) R. Wada and W. Yoshida, *Epigenomes*, 2021, **5**.

29) L. Y. Zhao, J. Song, Y. Liu, C. X. Song and C. Yi, *Protein & cell*, 2020, **11**, 792–808.

Wataru Yoshida

Biography: Wataru Yoshida is a professor at the School of Bioscience and Biotechnology at Tokyo University of Technology (TUT). He received his Ph.D. in Engineering from Tokyo University of Agriculture and Technology (TUAT) in 2008. In 2008, he was accepted for the JSPS Research Fellowship for Young Scientists (PD). In 2011, he became an assistant professor at TUAT and moved to TUT as an assistant professor in 2013. He became a senior assistant professor in 2017, an associate professor in 2021, and a full-time professor in 2024 at TUT. His research group is interested in analyzing the role of epigenetic modifications in the formation of non-canonical structures and in biomolecular engineering to develop biosensing technologies for epigenetic modifications. He is the author or co-author of more than 60 papers published in international peer-reviewed journals.