

(様式4)

## 学 位 論 文 概 要

平成 30年 1月 10日

学位申請者

BAY, DANITYAH HABIBALLAH T 印

### 学位論文題目

G-quadruplex structures as transcriptional regulators in *Dele* and *Cdc6* CpG islands, and as targets for DNA methylation detection

---

### 学位論文の要旨

G-quadruplex (G4) is a DNA secondary structure that has been shown to play an important role in biological systems. In this study, the biological functions of selected G4 sequences were studied using luciferase reporter assay. Results have shown that the luciferase expression was activated when G4-forming sequences from the *Dele* and *Cdc6* genes have been cloned in reporter vectors carrying a minimal promoter and the luciferase gene, and also have been detected in experiments applying a promoterless reporter vector. Furthermore, the activation was decreased by the telomestatin derivative L1H1-7OTD that can bind to G4 structure and stabilize it. When *Dele*-F, *Dele*-R and *Cdc6* CpG islands sequence were cloned in the promoterless reporter vector, the luciferase expression was activated with *Dele*-F CGI sequence and was inhibited by *Cdc6*, indicating that G4 formation is significant in these sequences for the transcriptional regulation. In conclusion, *Dele* and *Cdc6* G4 DNAs individually possess enhancer and promotor function; however, when studied in more complex CpG islands *Dele* G4 also demonstrates promotor activity, whereas *Cdc6* G4 may possess a dual function of transcriptional regulation.

DNA methylation has been considered to be a potential biomarker for cancer diagnosis. Methylated G4-forming sequences of mouse *Dele* and human *DELE* were studied to analyze the effect of methylated G4 on transcription. Additionally, a new DNA methylation detection system was developed utilizing G4 and i-motif forming sequences, based on the hypothesis that methylated G4 and i-motif structures may arrest DNA polymerase activity during quantitative polymerase chain reaction (qPCR). The results showed negative correlation between PCR amplification efficiency and methylation levels in the G4 and i-motif forming sequences. Results have demonstrated that DNA methylation of the G4 and i-motif forming sequences can be detected by analyzing the amplification efficiency using qPCR.

Tokyo University of Technology  
Graduate School of Bionics

**G-quadruplex structures as transcriptional regulators  
in *Dele* and *Cdc6* CpG islands, and as targets for  
detecting DNA methylation**

DOCTOR THESIS

by

BAY, DANİYAH HABIBALLAH T

2018

Doctor of Engineering

## TABLE OF CONTENTS

<b>DEDICATION</b>	<b>5</b>
<b>ACKNOWLEDGMENT</b>	<b>6</b>
<b>CHAPTER 1. Introduction</b>	<b>7</b>
1.1 Background	7
1.1.1 DNA structure	7
1.1.2 G-quadruplex (G4) structure	9
1.1.3 G4 location	13
1.1.4 G4 function	14
1.1.5 DNA methylation	18
1.1.6 G4 and DNA methylation	21
1.1.7 I-motif structure	22
1.2 Objective	24
<b>CHAPTER 2. Identification of transcriptionally regulating G4 structures in <i>Dele</i> and <i>Cdc6</i> CpG islands</b>	<b>25</b>
2.1 Background and aims	25
2.2 Materials and methods	26
2.2.1 Selection of G4-forming sequences	26
2.2.2 Plasmid construction	29
2.2.3 Cell culture	33
2.2.4 Luciferase reporter assays	34
2.2.5 CD spectroscopy	36

2.3	Results	38
2.3.1	<i>Dele</i> and <i>Cdc6</i> G4-forming sequences activate reporter gene expression	38
2.3.2	<i>Dele</i> and <i>Cdc6</i> CGIs	43
2.3.3	G4 ligand suppresses the transcriptional activation of <i>Dele</i> and <i>Cdc6</i> G4 DNAs	46
2.3.4	CD spectroscopy analysis of <i>Dele</i> and <i>Cdc6</i> G4 structures	50
2.4	Summary and discussion	58
	<b>CHAPTER 3. Methylated G4 structures and transcription</b>	<b>59</b>
3.1	Background and aims	59
3.2	Materials and methods	60
3.2.1	Preparation of methylated G4 DNA-containing plasmids for luciferase reporter assays	60
3.2.2	Cell culture and luciferase reporter assays	61
3.3	Results	62
3.3.1	Confirmation of methylated DNA vectors	62
3.3.2	Effect of methylated G4 structures on transcription	63
3.4	Summary and discussion	65
	<b>CHAPTER 4. Utilizing G4 and i-motif-forming sequences to detect DNA methylation</b>	<b>66</b>
4.1	Background and aim	66
4.2	Materials and methods	67
4.2.1	Preparation of methylated G4 and i-motif DNA	67

4.2.2	Quantitative PCR analysis of methylated G4 and i-motif DNA	70
4.2.3	Detection limit of methylated G4 and i-motif DNA in qPCR analysis	70
4.3	Results	72
4.3.1	Template DNA preparation	72
4.3.2	Methylated G4 and i-motif sequences affect qPCR amplification efficiency	75
4.3.3	Detection limit analysis	79
4.4	Summary and discussion	82
	<b>CHAPTER 5. Conclusions and prospects</b>	<b>83</b>
	<b>BIBLIOGRAPHY</b>	<b>85</b>
	<b>PUBLISHED PAPERS</b>	<b>100</b>

## **DEDICATION**

I dedicate my dissertation work to my loving parents, for which my mere expression of gratitude likewise does not suffice. Thank you my intelligent and strong mother, Dr. Samiha Algari, who fought the desire to keep her daughter near, and bottled up her feelings all these years living faraway, for the sake of her daughter getting higher education. Thank you for guiding me whenever I felt lost, and praising me whenever I achieved success. Thank you my knowledgeable and generous father, Habiballah Gary Bay, who has always put his trust in me and encouraged me to continue my studies. I cannot express my special feeling of gratitude to the unequivocal support of my beloved sisters, Dua'a and Morooj, and my guardian brothers Mohammad, Abdullah and Emran, whom have always helped me and encouraged me through the bad times, and the good times. I also dedicate this work to my sweat hearted grandmother, Aziza, for all the wholehearted prayers and caring love. I give my words of appreciation to the love and support of my family and cousins throughout the process. I give special thanks to my best friend since university years in Saudi Arabia, Dr. Amani Alhibshi, who was my ultimate support for most of my scholarship years in Japan, and who was with me throughout building and fulfilling our dream of traveling and living in Tokyo.

## ACKNOWLEDGMENT

This doctoral thesis would not have been possible without the help, support, and patience of my supervisor, Associate Prof. Wataru Yoshida, whose encouragement, guidance, and support from the initial to the final level enabled me to develop an understanding of the subject not to mention his good advice and support that has been invaluable on both an academic and a personal level, for which I am extremely grateful. I owe my deepest gratitude to the President of Tokyo University of Technology Prof. Karube Isao for his kindness, support, and encouragement, and for accepting me in this beautiful university, to be part of his laboratory for my doctoral program. I would like to thank Dr. Ikebukuro for supporting the submission of my article, Dr. Akimoto for lending necessary appliances for some experiments, Dr. Yano, Dr. Yokoyama and Dr. Imamura for the constructive criticism about my thesis.

I would like to acknowledge the greatest help of King Abdullah Scholarship Program and Umm Al-Qura University for providing the necessary financial support for my living and study expenses in Japan. I am blessed by and grateful for my beloved country Saudi Arabia. I am also thankful to the cultural office staff of the Saudi Arabian embassy in Japan for their help and support, especially Dr. Esam Bukhari. Last, but by no means least, I offer my regards and blessings to my colleagues in the laboratory for all the help and support.

# CHAPTER 1. Introduction

## 1.1 Background

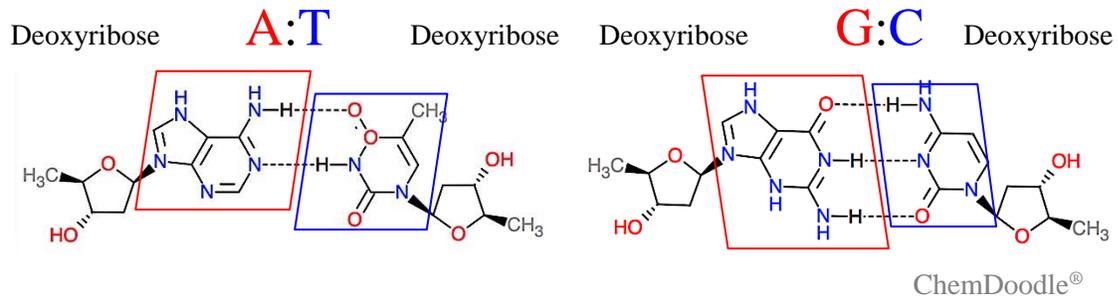
### 1.1.1 DNA structure

DNA structures predominantly take A-, B-, or Z-forms, and most DNAs adopt the right-handed duplex B form at neutral pH and physiological salt concentrations [1].

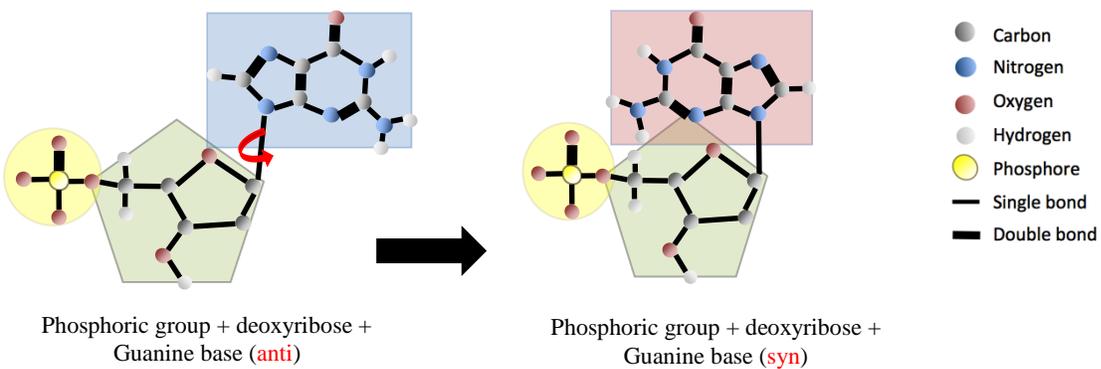
Duplex DNA comprises two strands in antiparallel 5' to 3' and 3' to 5' orientations, and these are coiled in the double helix structure proposed by Watson and Crick in 1953 [2].

Strands of double helical DNA are bound by purine–pyrimidine hydrogen bonds between adenine (A) and thymine (T), and between guanine (G) with cytosine (C). A:T base pairs form two hydrogen bonds and G:C base pairs for three (Figure 1.1), and whereas these are usually arrayed in the *anti* conformation, some bases can rotate by approximately 180° about the glycosidic bond to form the *syn* configuration (Figure 1.2) [3].

In addition to canonical base pairs, G:G and C:C interactions have been described and can lead to different DNA structures. The flexibility of double helix DNA allows folding into more complex three dimensional structures with three or four strands, forming secondary DNA structures, such as G-quadruplex and i-motif [4, 5, 6]. Following the human genome project, these alternative DNA structures have received increasing research attention.



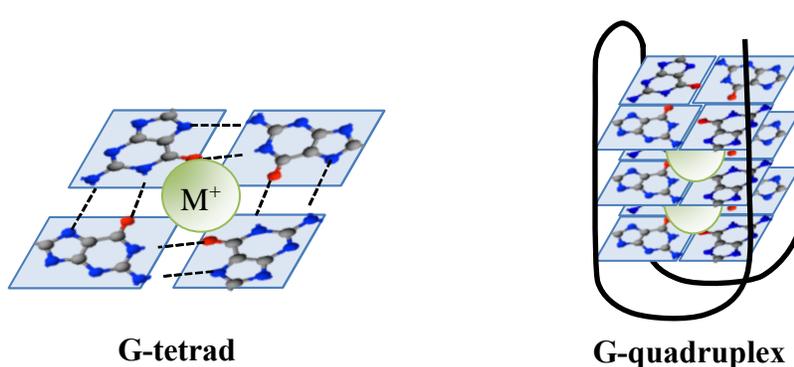
**Figure 1.1** Purine and pyrimidine bases are bound by 2 or 3 hydrogen bonds.



**Figure 1.2** *syn* and *anti* conformations of guanine bases about the glycosidic bond

### 1.1.2 G-quadruplex (G4) structures

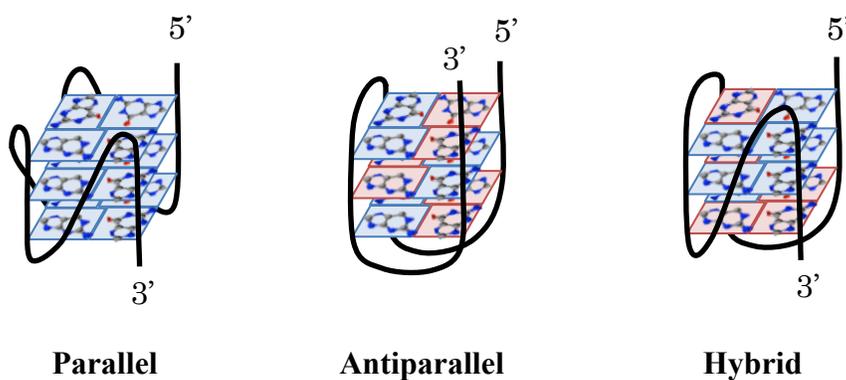
G4 DNA structure comprises two or more square planar guanine (G)-tetrads (Figure 1.3), which are composed of four G bases connected through Hoogsteen hydrogen bonds, involving inter- or intramolecular interactions of four DNA strands [5, 7, 8]. When G4 is folded, G-tetrads in  $\pi - \pi$  stacking conformations are usually stabilized by monovalent cations, such as potassium  $K^+$ , which fit between the two tetrads and is surrounded by 8 guanine  $O^6$  oxygen atoms (Figure 1.3) [9, 10, 11, 12].



**Figure 1.3** G-tetrad and G-quadruplex structures stabilized by monovalent cations

The topological diversity of G4 formations predominantly reflects variable orientations of DNA backbones and lengths of loop sequences connecting G runs, and G base alignments about glycosidic bonds in *anti* or *syn* conformations. In parallel-type G4 structures, all four backbone strands containing G runs are oriented in the same direction and all stacking G bases are in the *anti* conformation and are only rarely in the *syn* conformation (Figure 1.4) [13, 14]. Conversely, the term antiparallel-type refers to any other G4 structures that have one or two strands in different directions and base stacking

of both *anti* and *syn* configurations [15]. In recent studies, the term antiparallel-type has been used to describe G4 structures with two strands in the same direction and two strands in the opposite direction, and with base stacking in *anti/syn* or *syn/anti* conformations (Figure 1.4). In addition, hybrid-type or mixed-type G4 structures have one strand in the opposite direction to the other three strands and G base stacking of mixed *anti/syn* or *syn/anti* and *anti/anti* or *syn/syn* conformations (Figure 1.4) [16, 17].



**Figure 1.4** Parallel-, antiparallel-, and hybrid-type G4 structure orientations; guanine bases are colored red and blue for *syn* and *anti* conformations, respectively.

Nucleotides between pairs of G runs are involved in the formation of loops in G4 structures, and connect the strands as edge-wise lateral, diagonal, double-chain reversal, or propeller and V-shaped loops. Lateral loops connect neighboring antiparallel strands, whereas diagonal loops connect opposing antiparallel strands, propeller loops connect neighboring parallel strands, and V-shaped loops connect the corners of two G-tetrads that are missing a column [18, 19]. Critically, loops in G4 structures may act as targets for small ligand molecules [20].

Generally, G4 structures form on sequences with tracts of two or more G bases and loops of one or more nucleotides between the tracts. Accordingly, putative intrastrand G4-forming sequences can be predicted using computational algorithms, such as  $\mathbf{G}_x\mathbf{N}_{y1}\mathbf{G}_x\mathbf{N}_{y2}\mathbf{G}_x\mathbf{N}_{y3}\mathbf{G}_x$ , where  $x \geq 2$  and  $N_{yi}$  represents loop sequences [21], although these can vary in length, location, and numbers of strands involved in the folding process. Several studies report methods for detecting G4 forming sequences *in silico* and *in vitro* throughout the genome [22], and numerous intramolecular G4 structures have been identified. In addition, few approaches have been developed to identify motifs that contain G tracks on both strands of duplex DNA and form intermolecular G4 structures [23, 24].

Although G4 structures are thermodynamically more stable than their duplex counterparts [25], they are often temporary and are formed in equilibrium with duplexes, unless they are stabilized by factors such as ligands. Under pathological conditions such as cancer, this equilibrium is pushed in favor of folding or unfolding of the G4 structure, leading to overexpression or suppression of corresponding genes [26].

The vast topological diversity of G4 structures with differing folding patterns and loop lengths produces specific molecular properties, and these require dedicated investigations to target and manipulate structures of therapeutic value.

### 1.1.3 G4 location

Telomeric G4 structures have been studied in the human genome and in genomes of various other organisms, and have been shown to form telomeric repeats *in vitro* in NMR and X-ray crystal structure determinations [11]. Moreover, in genome-wide surveys, about 376,000 putative quadruplex sequences were identified in the human genome [27]. Among these, G-rich sequences are found upstream of proto-oncogene promoters, such as those of *c-MYC* [28], *c-KIT* [29], *VEGF* [14], *BCL-2* [30], and *RET* [31], and the *RAS* genes *HRAS* [32, 33] and *KRAS*. [34]. *In silico* analyses have also demonstrated that sequences with the potential to form G4 motifs are enriched in ribosomal DNA, immunoglobulin heavy-chain switch regions, and CpG islands (CGIs) [35, 36, 37, 38], suggesting widespread regulatory roles [39].

#### 1.1.4 G4 functions

G4 structures in telomeres have been shown to inhibit telomerase activities, which maintain telomere lengths and have been associated with more than 85% of cancers [40]. G4 structures also affect gene expression levels, and G4-forming sequences in promoter regions have been found to play functional roles in the suppression of proto-oncogenes [41, 42], such as *c-MYC* [28, 43, 44, 45, 46], *BCL-2* [47, 48], *VEGF* [49], and *RET* [31].

#### *G4 structures and transcriptional regulation*

A study using mutational analyses show that two neighboring DNAs upstream of the *HRAS* promoter repress transcription by forming G4 structures [32]. In addition, mutation of G4-forming sequences destabilizes G4 structures in the nuclease hypersensitivity element III<sub>1</sub> (NHEIII<sub>1</sub>) of the *c-MYC* promoter, resulting three-fold increases in basal transcription [28, 43]. In other studies, three individual G4 structures were identified in a GC-rich region upstream of the P1 promoter of *BCL-2* [48] and partial mutation of G to A in the ensuing sequences disrupted G4 structures and increased *bcl-2* transcription by two-fold [47].

G4 structures have also been associated with increased transcription. Specifically, mutation of a p32 G4-forming sequence, originally located in the P1 promoter of the *Bcl-2* gene, led to reduced transcriptional activity of mutated vectors in comparison with the native vectors [50], suggesting transcription activation by this G4 structure. Furthermore, in the insulin-linked polymorphic region (ILPR), G4 structures comprising two repeated consensus sequences [51, 52] activated transcription and single or double mutations reduced promoter activity [53, 54, 55]. In addition, the G4-forming *c-myb* GGA repeat

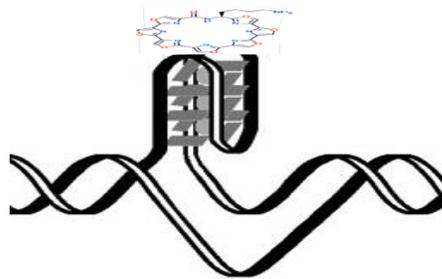
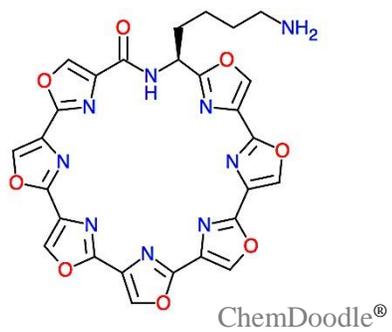
region reportedly acted as both transcriptional repressor and activator. Moreover, one or two deletions of (GGA)<sub>4</sub> motifs increased *c-myb* promoter activity, whereas deletion of all three regions eliminated promoter activity [56]. These regulatory activities of G4 structures suggest the potential of these structures as targets in therapeutic strategies for numerous human diseases, including cancer [57].

### ***Molecular interactions of G4 structures***

Targeting of G4 structures using proteins and small molecules has been shown to affect transcription in several studies. In particular, human cellular nucleic binding protein (CNBP) reportedly bound specifically to the single strand G-rich sequence of the *c-Myc* NHE III<sub>1</sub> region *in vitro*, and subsequently promoted the formation of G4 structures [43]. Moreover, overexpression of NM23-H2 strongly activated transcription of the *c-myc* gene by nearly 300% and interactions of CNBP and NM23-H2 also increased *c-Myc* transcription. These data indicate that increases in *c-Myc* expression following overexpression of CNBP is prompted by recruitment of NM23-H2 to the *c-Myc* NHE III<sub>1</sub> promoter due to their strong interaction suggesting the potential of CNBP as an anticancer target. In contrast, binding of the cationic porphyrin TMPyP4 to the *c-MYC* G4 repressed promoter activity [45]. Additionally, nucleolin has been identified as a *c-MYC* G4 binding protein that represses *c-MYC* expression [46] and *HRAS* transcription was dramatically inhibited by the G4 DNA ligand ATPD-1, which binds *HRAS* quadruplexes with high affinity. Accordingly, in the presence of abrogating point mutations of G4s, ATPD-1 repressed transcription by only 50% [32], whereas transcription of the *HRAS* promoter was activated by the zinc-finger protein MAZ, which lifted the transcriptional blockade caused

by the neighboring G4 at *hras-1* and *hras-2*. Other studies also demonstrate that targeting of G4 sequences using quindoline derivatives stabilized G4 structures, leading to reduced *Bcl-2* [47] and *VEGF* [49] transcription and decreased angiogenesis. In addition, the G4 ligands TMPyP4 and telomestatin reportedly stabilized G4 structures of the *RET* proto-oncogene promoter and reduced gene expression [31].

In a previous study, 1998 G4-forming sequences were identified in a mouse CGIs microarray using the fluorescent-labeled G4 ligand L1Cy5-7OTD [58]. Subsequent CD spectroscopy and DMS footprinting analyses confirmed the formation of 10 G4-forming sequences that were randomly selected. Thus, the aim of this study is to analyze the regulatory functions of these 10 G4 DNAs and to investigate the transcriptional effects of binding the telomestatin derivative G4 ligand L1H1-7OTD to G4 structures (Figure 1.5).



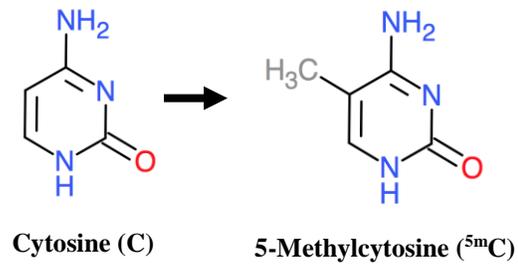
**Figure 1.5** Structure of the telomestatin derivative G4 ligand L1H1-7OTD that binds to G4 structures through stacking

### 1.1.5 DNA methylation

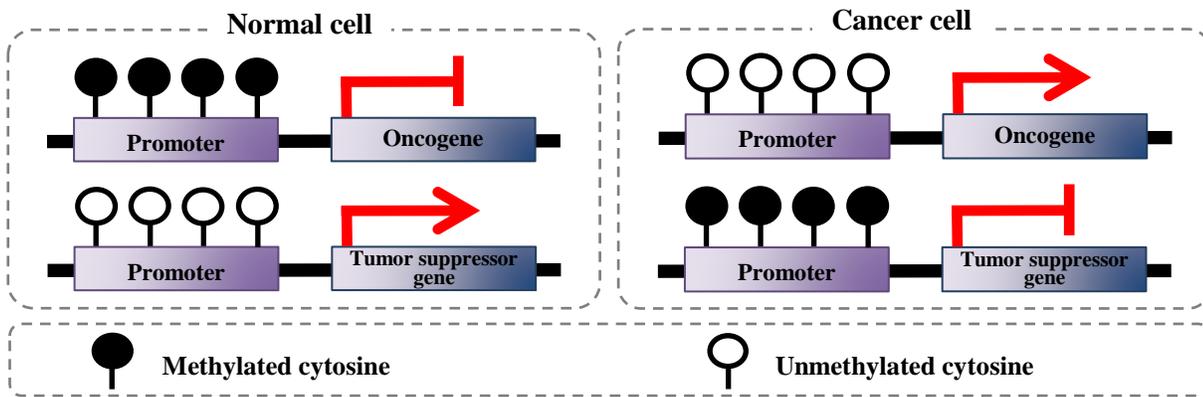
DNA methylation is an epigenetic mechanism that occurs by the attachment of a methyl group to a DNA base (Figure 1.6) [59]. Methylation of the O<sup>6</sup>-position of the G base (6 mG) has been associated with DNA damage and cancer [60, 61]. In addition, 6 mG reportedly induced structural conformational changes in G4 structures of human telomeres [62]. In mammalian genomes, methylation of cytosine is common in CGIs, and these are found in around 72% of promoters [63, 64] and have important roles in transcriptional regulation. Low rates of methylation, or hypomethylation, have been associated with overexpression of some oncogenes in cancer cells [65]. Conversely, hypermethylation in CGIs can inhibit the transcription of tumor suppressor genes, leading to tumorigenesis [66, 67] (Figure 1.7). These epigenetic observations have inspired several research groups to exploit DNA methylation patterns of some genes as biomarkers for cancer therapeutics [68, 69, 70].

Several methods have been developed to detect DNA methylation, such as sodium bisulfite method, and techniques using methylated DNA ligands have been widely reported. Sodium bisulfite modification involves the conversion of unmethylated C residues into uracil, which is then recognized as thymine in polymerase chain reaction (PCR) experiments. Concomitantly, methylated C residues remain unconverted and are amplified by specific primers. In these experiments, DNA is initially linearized using restriction enzymes, and is then denatured under alkaline conditions in the presence of sodium bisulfite/hydroquinone (pH 5.0) at 50°C for 16 or 40 h under mineral oil. The resulting solutions are then dialyzed at 4°C in large volumes of sodium acetate/hydroquinone, and unreacted bisulfite is

removed. Bisulfite-reacted DNA is then dried under vacuum and resuspended in buffer and ammonium acetate is added after incubation at room temperature for 10 min [66, 71]. Consequently, bisulfite assays are time consuming and expose the starting materials to low pH and high temperature, which can cause damage or loss [72, 73, 74]. Moreover, preparation of specific ligand molecules can be difficult and requires an additional step for ligand binding to methylated DNA [75, 76, 77]. Thus, more convenient methods for determining DNA methylation levels are eagerly awaited.



**Figure 1.6** Methylation of cytosine base



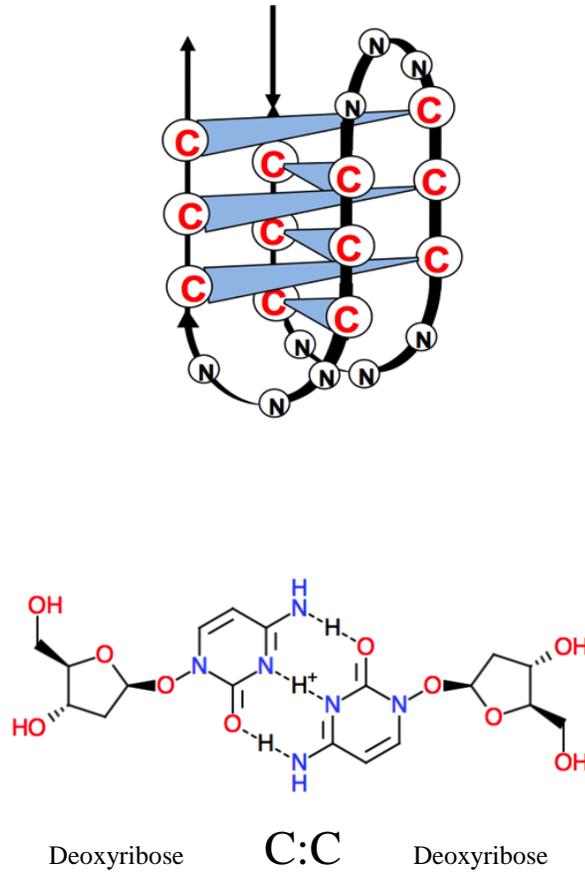
**Figure 1.7** Methylation of promoter DNA inhibits transcription

### **1.1.6 G4 and DNA methylation**

Stabilizing G4 structures that form near proto-oncogene promoters can influence gene expression [78] and several studies show increased stability of G4 structures following methylation. Specifically, C-5-methylation reportedly increased the thermal stability of G4 structure in P1 promoter of the anti-apoptotic factor *bcl-2* and inhibited extension reactions of DNA polymerase [79]. Although methylation of G4 structures also prevented PCR amplification of template DNA in a previous study [80], the effects of methylated G4 structures on transcription have not been investigated directly.

### 1.1.7 i-motif structures

Complementary strands of G-rich sequences that form G4 structures are cytosine (C)-rich sequences that form quadruplex structures known as intercalated (i)-motifs (Figure 1.8) [6]. These secondary DNA structures comprise two parallel-stranded DNA duplexes in an antiparallel orientation with intercalated C:C<sup>+</sup> base pairs. These i-motifs tend to fold and stabilize under acidic conditions [81, 82], and the resulting sensitivity to changes in pH can be exploited in nanotechnology applications. Studies are still limited about transcriptional functions of i-motif structures, but few have suggest that i-motif role may resemble that of G4s. The *VEGF* i-motif forming sequence [76] contains a binding site for the transcriptional factor Sp1, which plays a significant role in *VEGF* transcription [83]. Similar i-motifs have also been identified in association with *BCL-2* and *RET* genes [79, 31], and methylation of i-motif structures has been shown to increase base-pairing energies of proton-bound dimers of cytosine (C<sup>+</sup>·C) [84, 85], leading to greater stability of the structure.



**Figure 1.8** Schematic of an i-motif structure with the intercalating hydrogen bonding between cytosine bases; C:C<sup>+</sup>

## **1.2 Objective**

The main objective of this doctoral thesis is to analyze structural and functional properties of G4 structures, and to determine their effects on transcription. To this end, ten G4 forming sequences were randomly selected from CGIs of mouse genome and were cloned into luciferase reporter vectors (chapter 2). The effects of G4 formation on transcription were then analyzed using luciferase reporter assays after transfecting mammalian cells with wild-type and mutant G4 vectors. In these experiments, G4 structures were stabilized with L1H1-7OTD and the transcriptional effects were determined. Subsequently, circular dichroism (CD) spectroscopy analyses of G4 structures were performed in the presence and absence of L1H1-7OTD. In further experiments (chapter 3), the effects of methylated G4 structures on transcription were investigated using reporter assays. Finally, in hypothesis, PCR amplification efficiency decreases in the presence of methylated G4 and i-motif structures; therefore, methylated G4/i-motif forming sequences were used in a novel DNA methylation detection system based on amplification efficiency using qPCR (Chapter 4).

## **CHAPTER 2. Identification of transcriptionally regulating G4 structures in *Dele* and *Cdc6* CpG islands**

### **2.1 Background and aims**

G4 structures are secondary DNA structures with reported biological roles. In this study, 10 sequences were randomly selected from previously identified G4-forming sequences in mouse CpG islands, and their transcriptional effects were analyzed using luciferase reporter assays. Among these 10 G4 DNAs, the effects of *Dele* and *Cdc6* G4 DNAs were the strongest, and further reporter assays of *Dele* and *Cdc6* G4 DNA constructs were associated with corresponding topological data from CD experiments.

## 2.2 Materials and methods

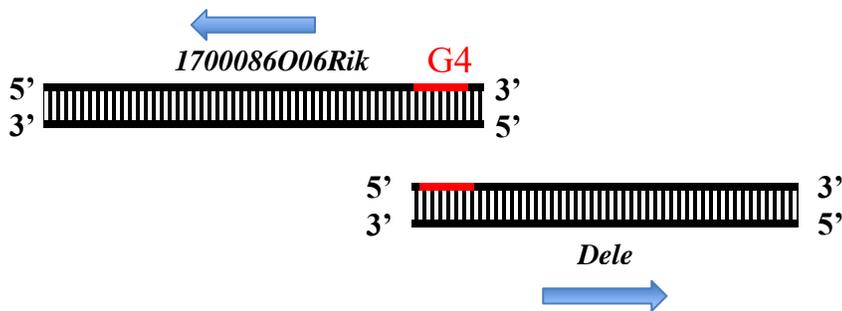
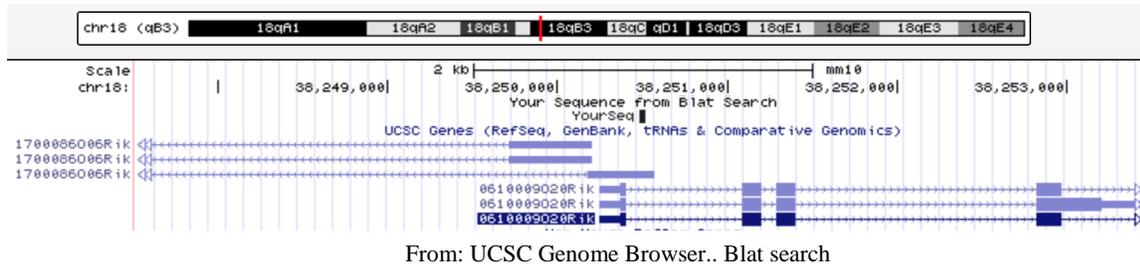
### 2.2.1 Selection of G4-forming sequences

To analyze transcriptional activities of G4 structures, wild-type and mutant constructs of the 10 selected G4 forming sequences were prepared and cloned into luciferase reporter vectors. These sequences were from *Jard2*, *Foxa2*, *Chd4*, *Med4*, *Bmi1*, *Wt1*, *Sp130*, *Cdc6*, and *Dele* genes (Table 2.1). Each sequence was constructed with additional bases on 3' ends to form *Sfi*I recognition site after cloning on the *Sfi*I site on the reporter vector.

#### *Dele* G4 DNA

The *Dele* G4-forming sequence was found on two divergently overlapping genes, *Dele* and *1700086O06Rik*, located on mouse chromosome 18 (Figure 2.1); therefore, the *Dele* G4 sequence was used in forward and reverse directions (*Dele*-F G4 and *Dele*-R G4). The gene *Dele* encodes the protein death ligand signal enhancer, which is essential for the induction of death receptor-mediated apoptosis via caspase activation. In contrast, the gene *1700086O06Rik* encodes an as yet uncharacterized protein.

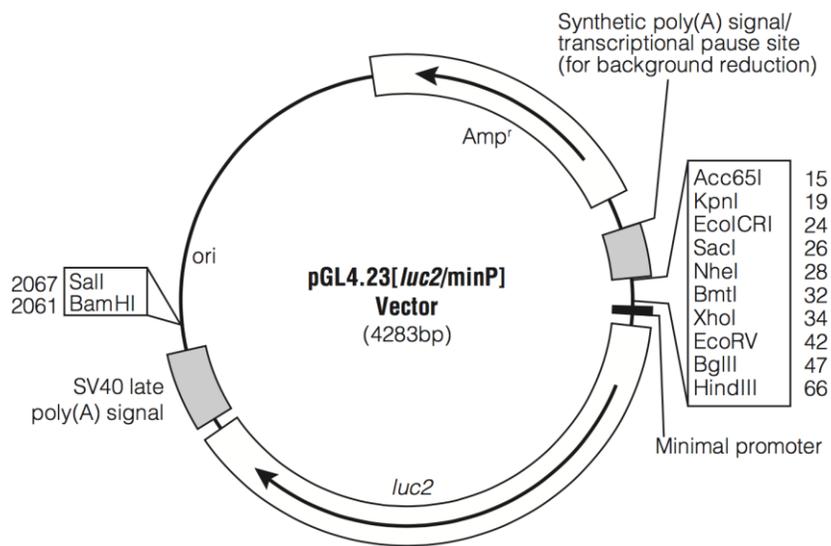




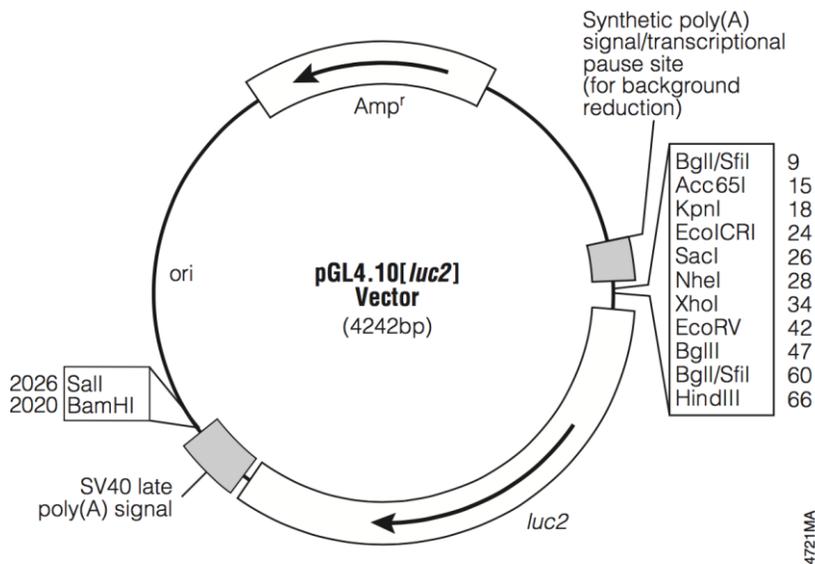
**Figure 2.1** G4-forming sequence, located on the divergently overlapping *Dele* and *1700086006Rik* genes, was used as *Dele-F* in the forward direction and as *Dele-R* in the reverse direction.

### 2.2.2 Plasmid construction

Wild-type and mutant G4 DNAs were cloned into the *Sfi*I sites of the minimal promoter containing vector pGL4.23 [luc2/minP] (Figure 2.2) and the promoterless vector pGL4.10 [luc2] (Figure 2.3; Promega Corporation, Madison, WI, USA). Ligated plasmids were then amplified by the transformation of *E. coli* DH5 $\alpha$  competent cells (TOYOBO, Osaka, Japan). Next, plasmids were extracted using PureYield Plasmid Miniprep System (Promega Corporation, Madison, WI, USA). In order to eliminate the least possibility of the plasmids to get methylated, which may consequently interfere with transcription of G4 regions, *E. coli* HST04 *dam*-/*dcm*- competent cells (Takara, Tokyo, Japan) was transformed by the plasmids to prepare DNA methylation-free plasmids.). All plasmids were sequenced to confirm target sequences using a 3730x1 DNA analyzer (Thermo Fisher Scientific, Waltham, MA, USA).



**Figure 2.2** Minimal promoter containing vector pGL4.23



**Figure 2.3** Promoterless vector pGL4.10

*Dele* and *Cdc6* CGIs were amplified from C57BL/6 mouse genomic DNA, via PCR, using the primers shown in Table 2.2. PCR products were then purified using Wizard SV Gels and PCR Clean-Up System kit (Promega Corporation, Madison, WI, USA) and were then digested using *Sfi*I (NEB, Ipswich, MA, USA). Products were then cloned into the *Sfi*I site of pGL4.10 [luc2] and the plasmids were prepared as described above. Successfully ligated vectors were discriminated from the unmodified pGL4 luciferase reporter vector according to the absence of the *Sfi*I cleavage site after insertion of the G4 sequence. Ligation was subsequently confirmed by digesting 0.5 µg of the ligated product with 10 U of *Sfi*I in a reaction volume of 10 µL at 50°C for 1 h. Products were then electrophoresed on 1% agarose gels, and ligation was confirmed according to migration of DNA bands. Positively screened vectors were then sequenced by MACROGEN, Japan. To construct mutant-type G4 vectors, site-directed mutagenesis was performed using KOD - plus- mutagenesis kits (TOYOBO, Osaka, Japan) according to the manufacturer's protocol with the primers for mutants shown in table 2.2.

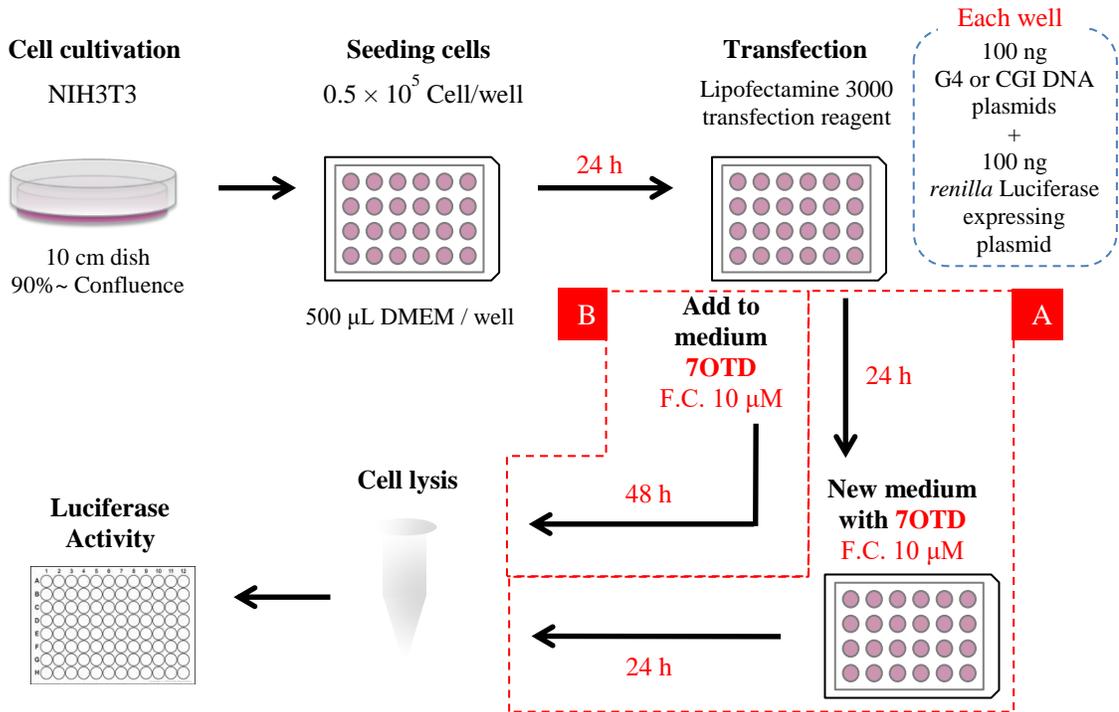
**Table 2.2** Primer sequences used for CGI vector construction and site-directed mutagenesis

Mutation sites are shown in bold.

Primer		Sequences (5'-3')
<i>Dele</i> -F CGI	Forward	ATTGGCCTAACTGGCCAGGGTGCTCTAGGTTACCA
	Reverse	ATTGGCCGCCGAGGCCTCCCCTTGGACCTAAGCTCT
<i>Dele</i> -R CGI	Forward	ATTGGCCTAACTGGCCTCCCCTTGGACCTAAGCTCT
	Reverse	ATTGGCCGCCGAGGCCTAGGGTGCTCTAGGTTACCA
<i>Cdc5</i> CGI	Forward	ATTGGCCTAACTGGCCATGAGCAAAGGTAGCCCAGT
	Reverse	ATTGGCCGCCGAGGCCTGCTCAAACTAGCCAGCA
<i>Dele</i> CGI MT	Forward	CGGGACAGAGGGAGCGAGG
	Reverse	<b>AAA</b> TTCCCAGATCTAAGCCCACCCACT
<i>Cdc5</i> CGI MT	Forward	TGGAGGACAAAGTAGAAATAAAAATACG
	Reverse	<b>AAA</b> AGCCTCCCCACCGTTGCC

### 2.2.3 Cell culture

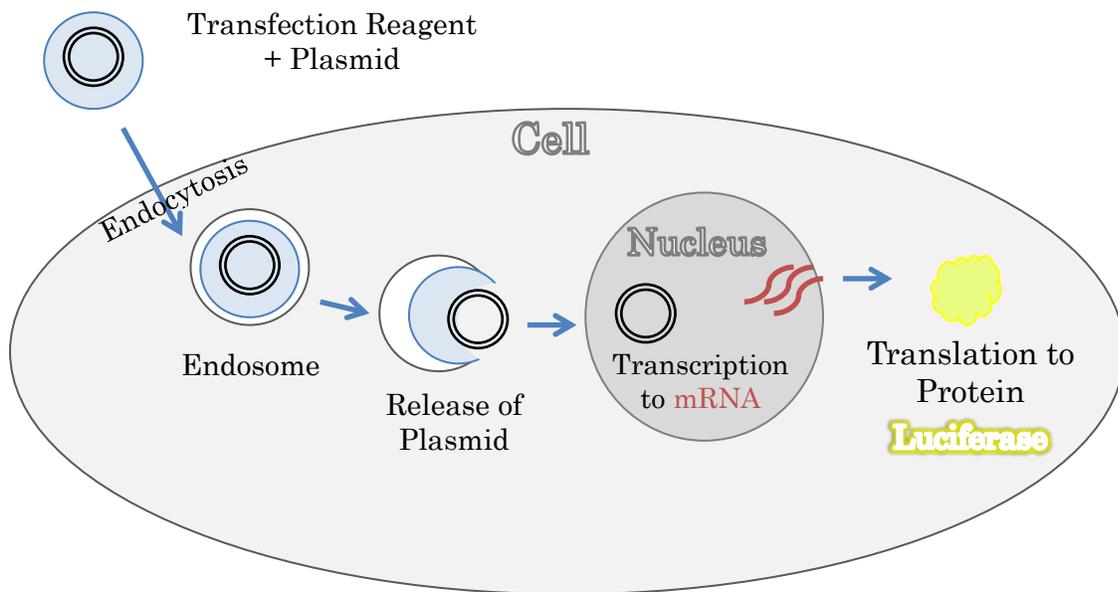
NIH3T3 cells were cultured in dulbecco modified eagle medium (DMEM) medium (Sigma-Aldrich, St. Louis, Missouri, USA) containing 10% fetal bovine serum (Sigma-Aldrich, St. Louis, Missouri, USA) and  $1 \times$  Penicillin-Streptomycin-L-Glutamine Solution (Wako, Tokyo, Japan) at 37°C in 5% CO<sub>2</sub>. Cells were passaged at about 90% confluence after 2–3-days culture (Figure 2.4).



**Figure 2.4** Transfection methods for reporter assays with and without the ligand. Ligand was added after one day and on the day of transfection in methods [A] and [B], respectively.

## 2.2.4 Luciferase reporter assays

NIH3T3 cells were transfected with 100 ng of the firefly luciferase reporter vector and 100 ng of the Renilla luciferase control vector using the transfection reagent Lipofectamine 3000 (Thermo Fisher Scientific, Waltham, MA, USA) according to the manufacturer's protocol. After 48 h, cells were lysed and luciferase activities were measured using Dual-Luciferase Reporter Assay System kit (Promega Corporation, Madison, WI, USA) and SPARK 10M microplate reader (TECAN, Männedorf, Switzerland). Firefly luciferase expression levels in cells were determined as the ratio of firefly to Renilla luciferase activity and expression levels were normalized to those of pGL4.10 or pGL4.23 luciferase reporter vectors (Figure 2.4, 2.5). All reporter assays were performed in triplicate.



**Figure 2.5** Schematic of cell transfection and luciferase protein expression

The transcriptional effects of the G4 ligand were investigated using two methods. In method A (Figure 2.4 A), 100 ng of firefly luciferase reporter vector and 100 ng of Renilla control vector were transfected into NIH3T3 cells using Lipofectamine 3000. After 24 h, culture media were refreshed and cells were exposed to 10  $\mu$ M L1H1-7OTD for a further 24 h prior to determining luciferase activity as described above. In method B (Figure 2.4 B), culture media were changed and cells were exposed to 10  $\mu$ M L1H1-7OTD prior to transfection with 100 ng of firefly luciferase reporter vector and 100 ng of the Renilla control vector in the presence of 10  $\mu$ M L1H1-7OTD using Lipofectamine 3000. Luciferase expression levels were determined in the presence or absence of L1H1-7OTD and were normalized to those of the pGL4.23 luciferase reporter vector in the presence or absence of L1H1-7OTD. All reporter assays were performed in triplicate.

### 2.2.5 CD spectroscopy

Wild-type and mutant *Dele* and *Cdc6* G4 oligonucleotides were purchased from Macrogen, South Korea (Table 2.3) and prepared in distilled water to create stock solutions (100  $\mu$ M). Oligonucleotides were then diluted to 15  $\mu$ M in TK buffer containing 50 mM Tris-HCl and 100 mM KCl (pH 7.5) prior to use in experiments. Oligonucleotides were denatured at 95°C for 3 min, were cooled to room temperature for 30 min, and were then diluted to 10  $\mu$ M and incubated in the presence or absence of 10  $\mu$ M L1H1-7OTD for 10 min. CD spectra were then measured using a J-1500 CD Spectrometer (JASCO, Tokyo, Japan) at 220–320 nm using a 1 mm path-length cuvette in 5°C increments from 20°C to 95°C. Baseline spectra were corrected for signal contributions of the buffer with and without the G4 ligand. CD melting analyses were performed with wild-type *Dele* and *Cdc6* G4 DNAs, and molar ellipticities were measured concomitantly in 1°C intervals at 264 nm and 262 nm, respectively.

To determine  $T_m$  values, the molar ellipticity at 25°C was set as 100%, and the molar ellipticity at 95°C was set as 0%. Molar ellipticities were normalized by curve fitting using GhraphPad Prism7 software, and  $T_m$  values were calculated as the temperatures corresponding to 50% normalized molar ellipticity.

**Table 2.3** G4 DNA sequences used in CD Spectroscopy

Mutation sites are shown in bold.

Name	Sequences (5'-3')
Mouse <i>Dele</i> G4	GGGTGGGCTTAGATCTGGGAAGGGCGGG
Mouse <i>Dele</i> G4 MT	GGGTGGGCTTAGATCTGGGA <b>ATTC</b> GGG
Mouse <i>Cdc6</i> G4	GGGGAGGCTGGGTGGAGG
Mouse <i>Cdc6</i> G4 MT	GGGGAGGCT <b>TTTT</b> GGAGG

## 2.3 Results

### 2.3.1 *Dele* and *Cdc6* G4-forming sequences activate reporter gene expression

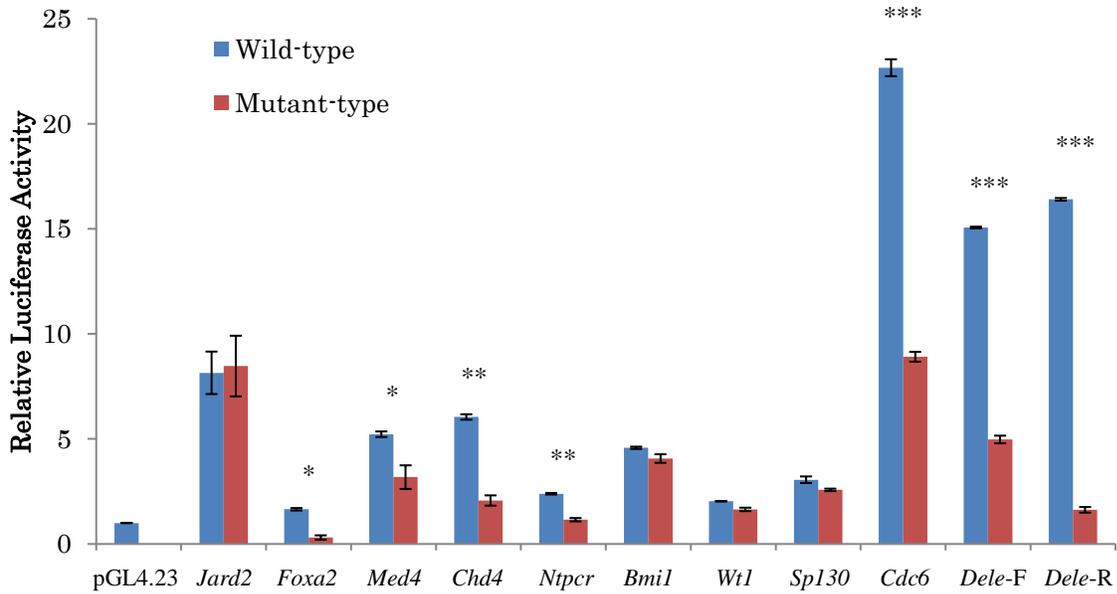
Plasmids for reporter assay experiments were constructed using G4 DNA sequences from *Jard2*, *Foxa2*, *Med4*, *Chd4*, *Ntpcr*, *Bmi1*, *Wt1*, *Sp130*, *Cdc6*, and *Dele* genes and luciferase reporter vectors with a minimal promoter containing a TATA-box promoter element. Plasmid constructs were confirmed using gel shift assays and sequencing of the reporter vector.

Reporter vectors were transfected into NIH3T3 cells and luciferase gene expression was measured after 48 h culture. Luciferase activities were normalized to those of the Renilla luciferase expressing vector pGL4.74 and all measurements were performed in triplicate. These experiments (Figure 2.6) demonstrated greater expression values for *Cdc6*, *Dele-F*, and *Dele-R* G4 DNAs than for the other G4 DNAs.

To determine the transcriptional effects of G4 structures, mutant vectors were designed by replacing single guanine run with thymine. These guanines were reported to be strongly associated with G4 formation, by DMS footprinting results in a previous study [58].

Comparisons of expression levels of G4-forming sequences and mutant sequences demonstrated significant differences for *Foxa2*, *Med4*, *Chd4*, *Ntpcr*, *Cdc6*, *Dele-F* and, *Dele-R*, but not for *Jard2*, *Bmi1*, *Wt1*, and *Sp130* (Figure 2.6). These results indicate that *Foxa2*, *Med4*, *Chd4*, *Ntpcr*, *Cdc6*, *Dele-F*, and *Dele-R* G4 DNA sequences may possess an enhancer activity on the luciferase reporter vector, suggesting the formation of specific G4 structures that may enhance the transcription. In particular, wild-type *Cdc6* and *Dele* G4

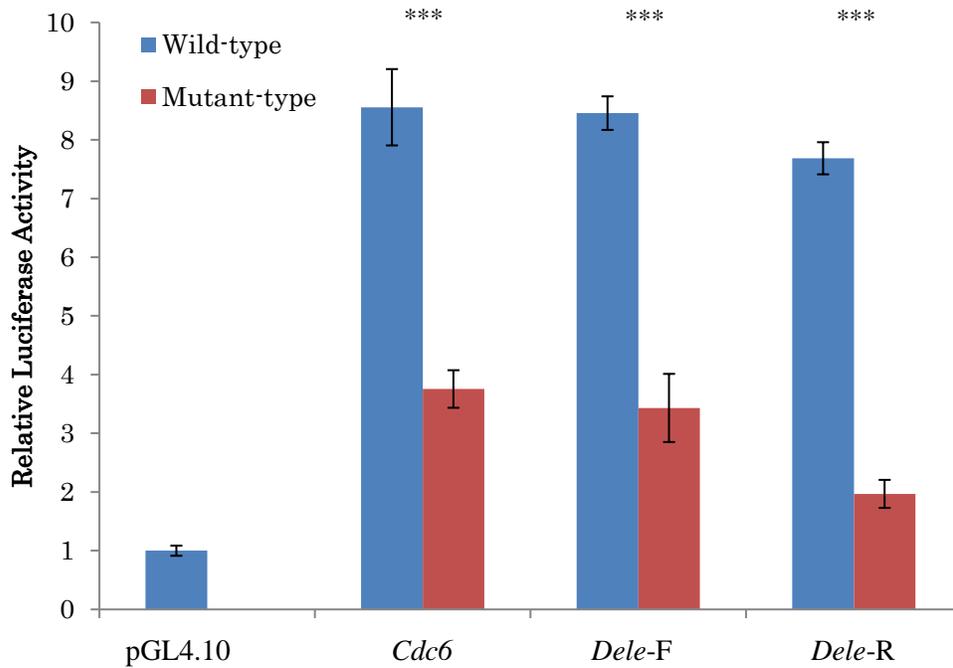
DNAs have highly activated the transcription with a high significant difference ( $P < 0.0001$ ) compared with corresponding mutants, while the activity was with lower significance in *Foxa2* and *Med2* ( $P < 0.01$ ) and *Chd4* and *Ntpcr* ( $P < 0.001$ ); therefore, *Cdc6* and *Dele* G4 DNAs were selected to perform further analysis.



**Figure 2.6** Luciferase reporter assays of enhancer activities of G4 DNA sequences

G4-forming sequences were cloned into the pGL4.23 vector, which has a minimal promoter. Blue bars represent wild-type constructs and red bars represent the mutant constructs. Luciferase activities relative to the pGL4.23 vector are shown as means  $\pm$  standard deviations (SD; n = 3); \*P < 0.01, \*\*P < 0.001, \*\*\*P < 0.0001

In addition to interactions between enhancers and promoters, promotor–promotor interactions are considered feasible [86], indicating that some promoters can also act as enhancers for other gene promoters. Therefore, to investigate whether *Dele* and *Cdc6* G4 DNAs possess promoter activities, reporter assays were performed using promoterless vectors. These experiments demonstrated that *Dele-F*, *Dele-R*, and *Cdc6* G4 DNAs activated luciferase expression (Figure 2.7) and that protein expression is decreased by thymine mutations in the G4 region. These results indicated that *Dele* and *Cdc6* G4 DNAs play regulatory roles as transcriptional promoters and enhancers, and suggest that the formation of secondary structures is essential for these effects.



**Figure 2.7** Reporter assays of promoter activities of *Cdc6*, *Dele-F*, and *Dele-R* G4 DNAs. G4-forming sequences were cloned into the promoterless vector pGL4.10. Blue bars represent wild-type and red bars represent mutant vectors. Luciferase activities are presented as means  $\pm$  SD relative to the pGL4.10 vector (n = 3) and differences were identified using t-test; \*\*\*P < 0 .0001

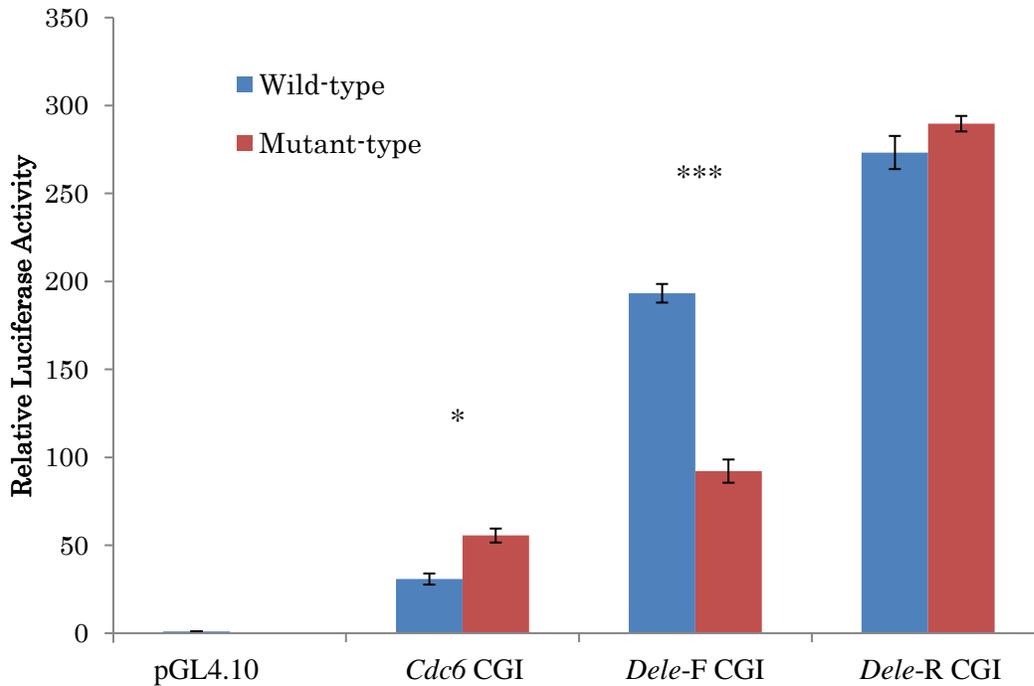
### 2.3.2 *Dele* and *Cdc6* CGI

To assess the functions of *Dele* and *Cdc6* G4 DNAs in CGIs that may contain more regulatory elements, a 696-bp DNA fragment containing a 475-bp *Cdc6* CGI and a 555-bp DNA fragment containing a 477-bp *Dele* CGI were cloned into the promoterless vector and transcriptional activities were then determined using reporter assays. These experiments showed greater than 10-fold increases in overall luciferase activities of vectors containing *Cdc6*, *Dele-F*, and *Dele-R* CGI DNAs (Figure 2.8). Moreover, luciferase activities were higher than those of individually cloned G4 sequence vectors, suggesting that G4 structures interact with cis-regulatory elements within CGIs that activate transcription.

In contrast with the vector containing the *Cdc6* G4-forming sequence, mutant *Cdc6* CGI DNA had higher transcriptional activity than its wild-type counterpart. These observations suggest dual transcriptional functions of *Cdc6* G4, as shown previously for *c-myc* G4 [43], which acted as an activator and a suppressor depending on the presence of transcriptional factors.

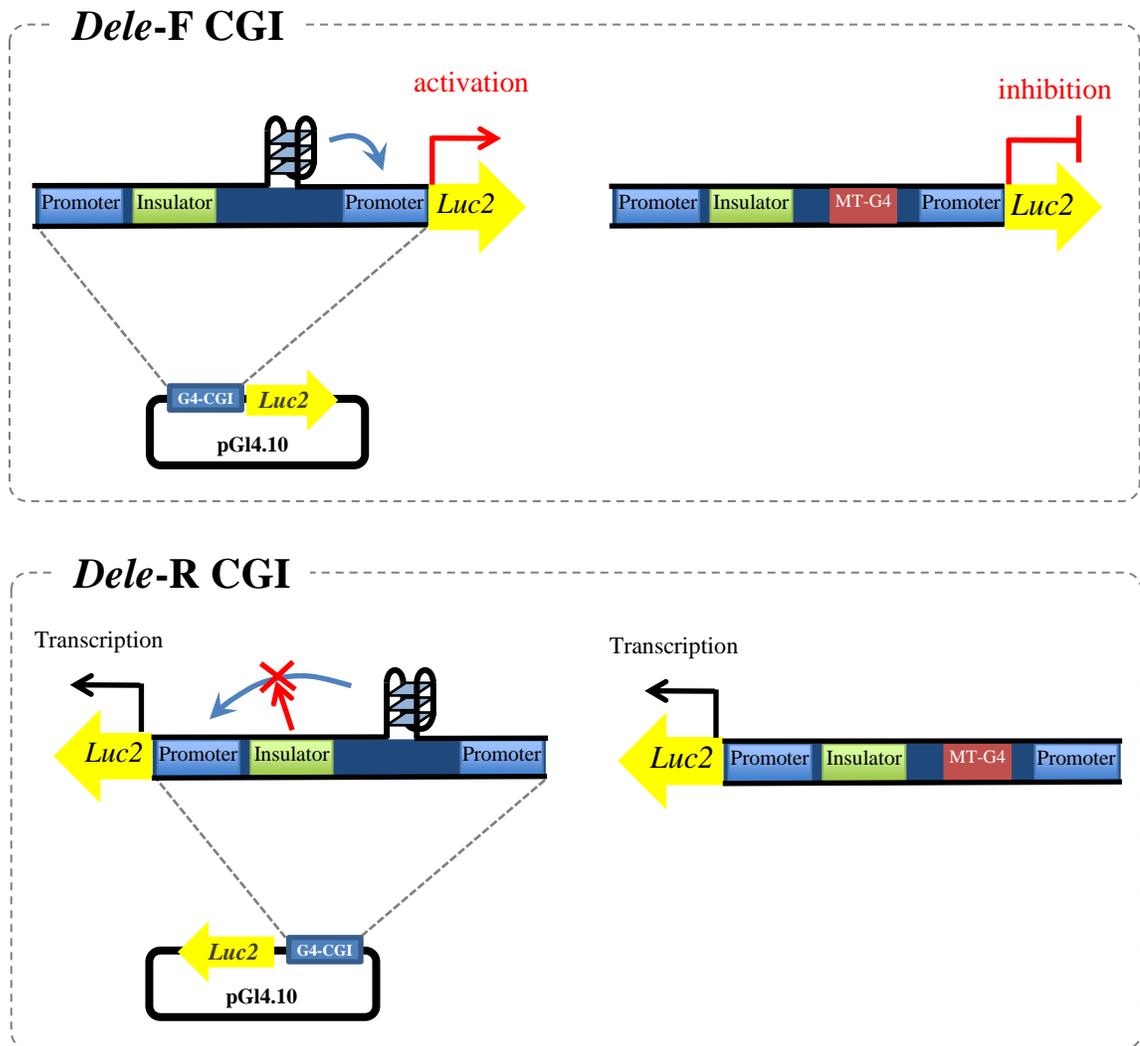
The transcriptional activity of mutant *Dele-F* CGI DNA was significantly ( $P < 0.0001$ ) decreased compared with that of the wild-type, indicating a high transcriptional activity of the G4 structure within this CGI sequence. In contrast, the transcriptional activity of the *Dele-R* CGI mutant was similar to that of its wild-type counterpart. The *Dele* CGI is located on the divergently overlapping genes *Dele* and *1700086O06Rik*, suggesting the presence of at least two regulatory sequences. Results indicate that the *Dele* CGI sequence may contain a promoter for *1700086O06Rik*, an insulator, and a promoter for

*Dele* (Figure 2.9). According to a previous study, the insulator sequence may have blocking activity that prevents promoter–enhancer interactions [87], likely preventing detection of the enhancer activity of the *Dele* G4 on the *1700086O06Rik* promoter.



**Figure 2.8** Reporter assays of transcriptional effects of *Cdc6*, *Dele-F*, and *Dele-R* G4 DNAs in CGI sequences

Luciferase activities were normalized to that of pGL4.10 and are presented as means  $\pm$  SD (n = 3). Differences between wild-type and mutant constructs were identified using t-tests; \*P < 0.01, \*\*\*P < 0.0001.



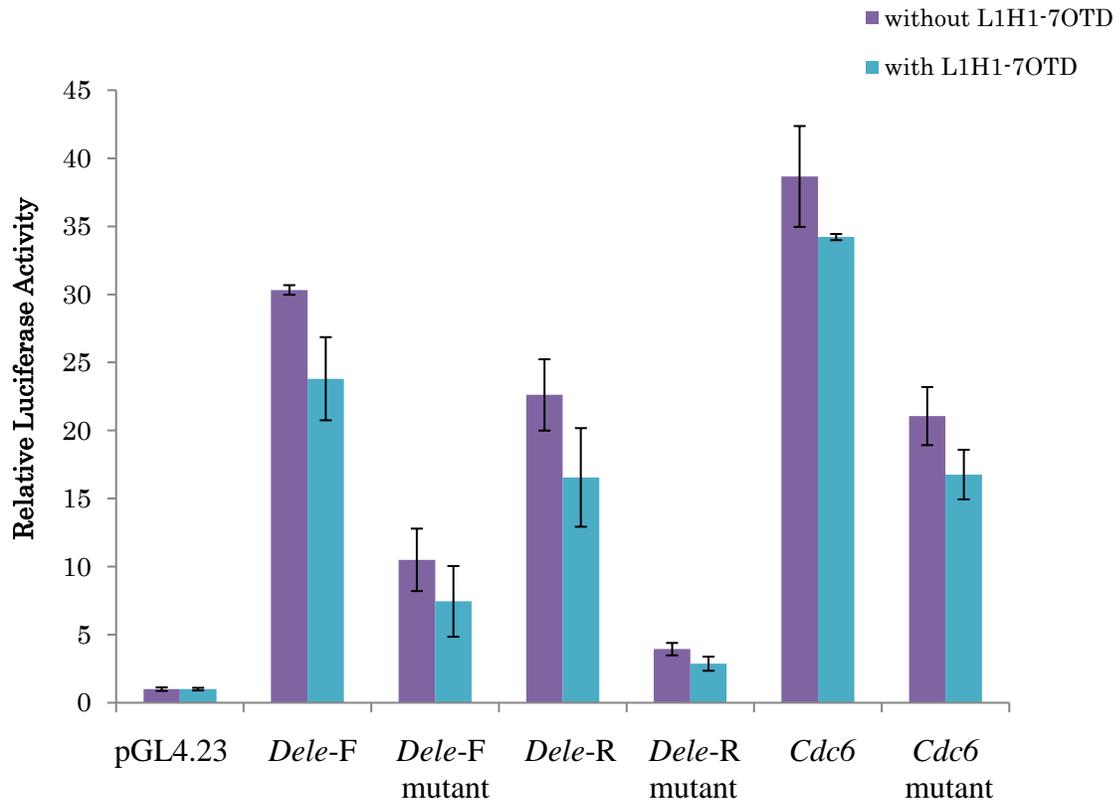
**Figure 2.9** Schematic of the role of the *Dele* G4 region within the CGI after cloning onto the vector in forward (*Dele-F CGI*) and reverse (*Dele-R CGI*) directions

### 2.3.3 G4 ligand suppresses the transcriptional activation of *Dele* and *Cdc6* G4 DNAs

G4 binding ligands have been developed in several studies of G4 structure–function relationships. These small molecules can distort, stabilize or induce structural changes in secondary DNA structures. Herein, the effects of the telomestatin derivative L1H1-7OTD on the transcriptional activity of *Dele* and *Cdc6* G4 DNAs were determined after binding to G4 structures. This ligand was previously shown to bind the top G-tetrad structure through  $\pi$ -stacking and electrostatic interactions [88, 89].

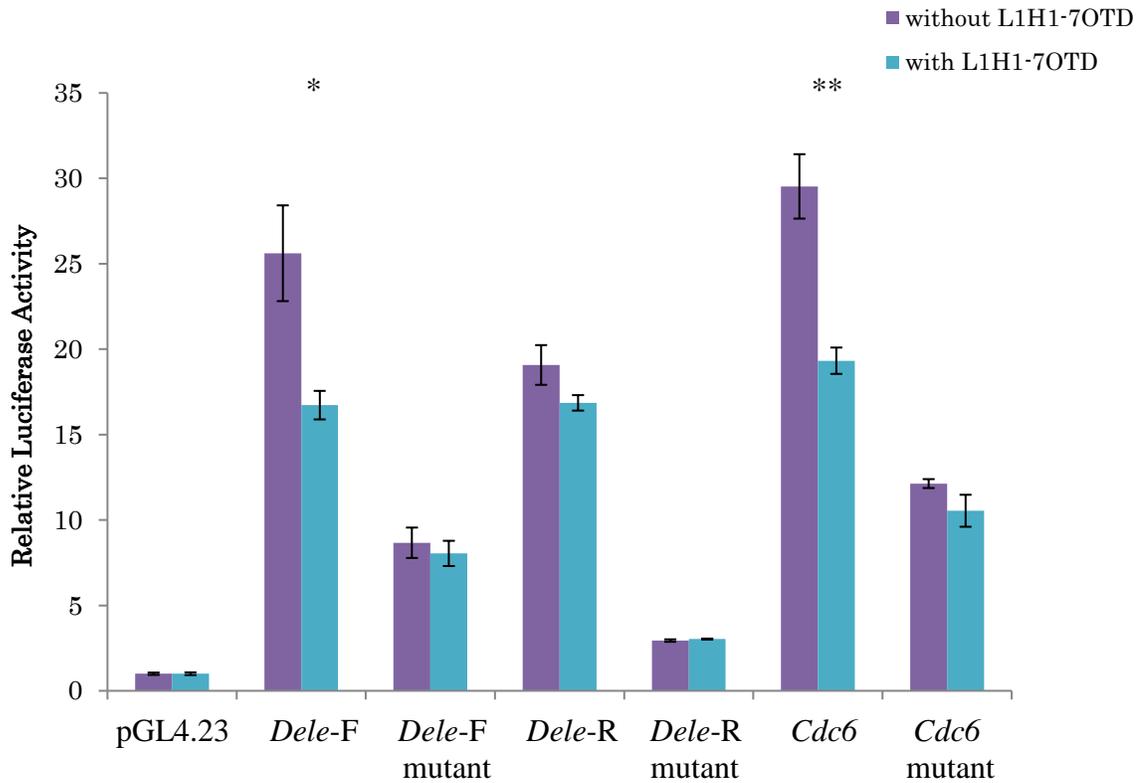
In reporter assays using method A, *Dele-F*, *Dele-R*, and *Cdc6* G4 DNAs were inserted into reporter vectors with a minimal promoter, and were then transfected into NIH3T3 cells in the presence of L1H1-7OTD. Subsequently, media were refreshed and cells were exposed to the test ligand at 1 day after transfection and were cultivated for another day prior to measuring luciferase activities. In these assays, data for ligand-treated and -untreated cells were normalized to those of ligand-treated and -untreated controls, respectively, and no significant differences in luciferase activity were identified (Figure 2.10). Because these observations suggest poor membrane permeability of the G4 ligand, subsequent experiments were performed using method B, in which vectors were mixed with the G4 ligand prior to transfection into NIH3T3 cells in ligand-supplemented medium. The ensuing data demonstrated 35% to 37% inhibition of luciferase expression from wild-type G4 ligand-treated *Dele-F* and *Cdc6* G4 DNA constructs, but no significant differences in the presence of transfected mutant vectors (Figure 2.11). These observations of inhibition suggest that the ligand may have induced changes in G4 structures; however, no structural changes were observed in CD spectroscopy analyses (as explained later),

indicating that inhibitory activities of the ligand follow binding to the top of the G4 structure, may have interfered with interactions between G4 structures and respective transcription factors by influencing dynamics, charges, and steric conditions.



**Figure 2.10** Reporter assays showing the effects of L1H1-7OTD on *Cdc6*, *Dele-F*, and *Dele-R* G4 DNAs using method A (Figure 2.4).

Violet bars represent relative luciferase activities in the absence of L1H1-7OTD, and blue bars represent those in the presence of L1H1-7OTD. Luciferase activities relative to that of the pGL4.23 vector are presented as means  $\pm$  SD ( $n = 3$ ). No significant effects of the ligand were identified using t-tests.



**Figure 2.10** Reporter assays showing the effects of L1H1-7OTD on *Cdc6*, *Dele-F*, and *Dele-R* G4 DNAs using method B (Figure 2.4)

Violet bars represent relative luciferase activities in the absence of L1H1-7OTD, and blue bars represent those in the presence of L1H1-7OTD. Luciferase activities relative to that of the pGL4.23 vector are presented as means  $\pm$  SD (n = 3). The ligand L1H1-7OTD significantly reduced luciferase activities of *Cdc6* and *Dele-F* constructs; \*P < 0.01, \*\*P < 0.001

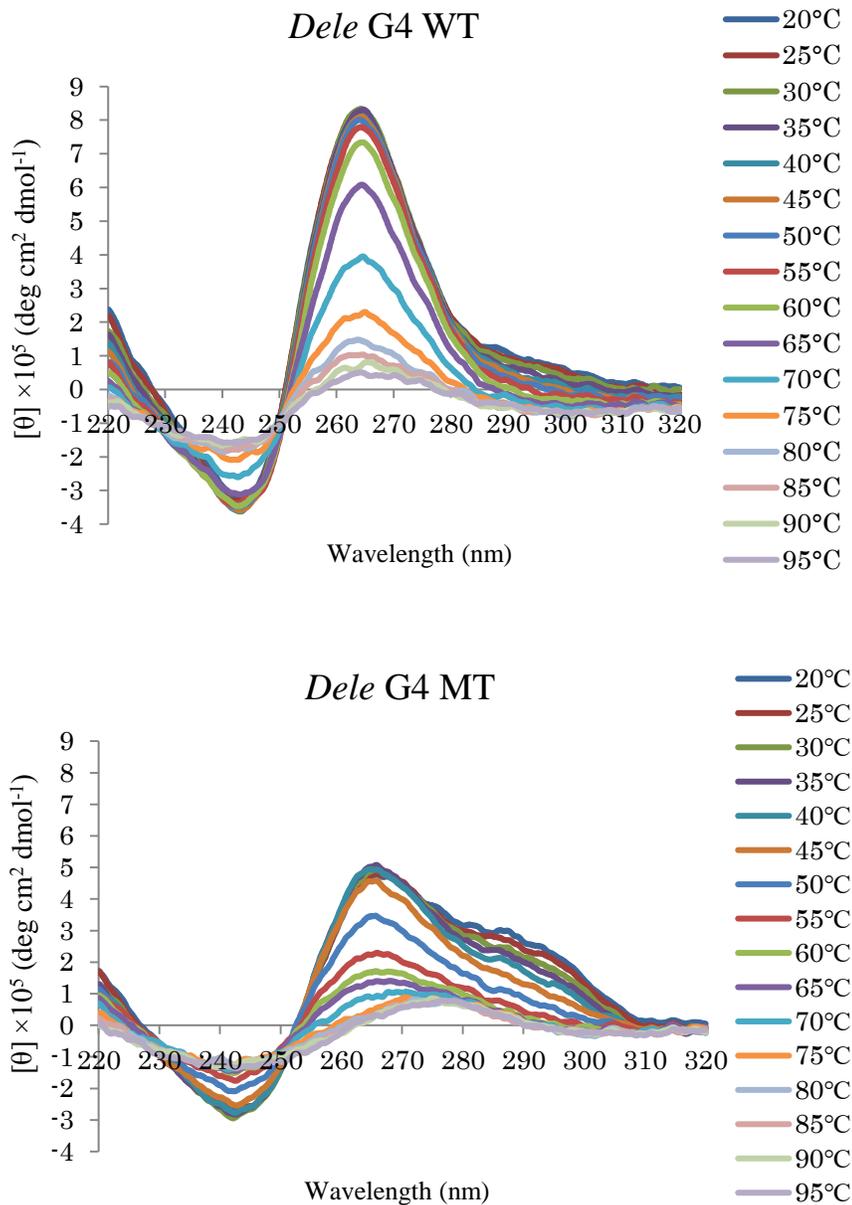
### 2.3.4 CD spectroscopy analyses of *Dele* and *Cdc6* G4 structures

To confirm the formation of G4 structures, CD spectral analyses of wild-type and mutant *Dele* (Figures 2.12) and *Cdc6* G4 DNAs (Figures 2.13) was performed. Subsequently, positive cotton effects were observed in wild-type *Dele* G4 DNA and *Cdc6* G4 DNA at around 264 and 262 nm, respectively, and negative cotton effects in wild-type *Dele* G4 DNA and *Cdc6* G4 DNA at around 242 and 240 nm, respectively. These data indicate the formation of parallel type G4 structures, whereas CD analyses of mutant G4 DNAs showed disrupted structures in comparison with wild-type structures.

CD spectra changed little in the presence of the G4 ligand L1H1-7OTD, indicating that the formation of parallel type G4 structures in wild-type *Dele* and *Cdc6* G4 DNAs is not disturbed by ligand binding (Figures 2.14, 2.15) [15]. CD spectra of mutant sequences (mutation of a single G-run to a T-run) show low molar ellipticity at around 260 nm and a shoulder between 280 and 300 nm, indicating that a transient secondary structures may form but not related to characteristic G4 structures (Figure 2.16).

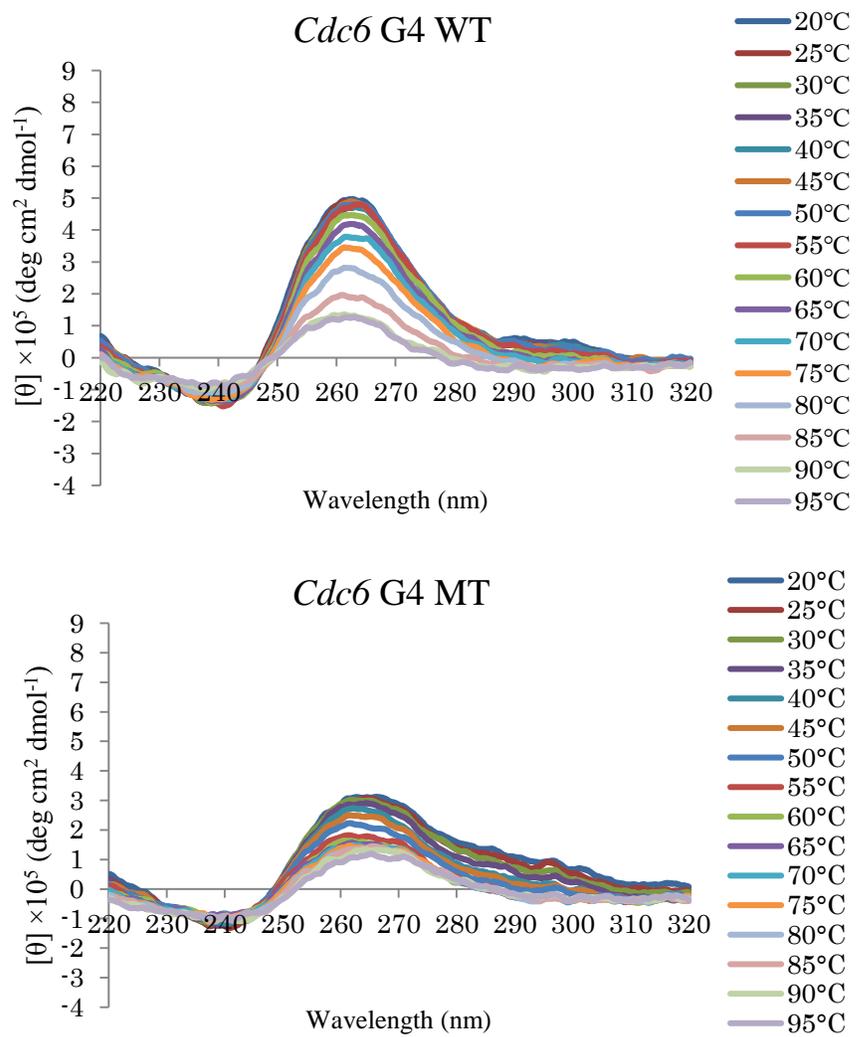
In CD melting analyses,  $T_m$  values (Figure 2.17) were 70°C for *Dele* G4 DNA at 264 nm and 77°C for *Cdc6* G4 DNA at 262 nm in the absence of the G4 ligand. These  $T_m$  values increased to 72°C for *Dele* G4 DNA and to 81°C for *Cdc6* G4 DNA in the presence of the G4 ligand, indicating that ligand binding to the G4 structure of *Dele* and *Cdc6* increases the stability of G4s under the present experimental conditions. These findings are in agreement with a NMR study on the structurally similar macrocyclic compound L2H2-6M(2)OTD, which binds to the top G-tetrad structure via  $\pi$ -stacking and electrostatic interactions without changing the folding topology [90]. Additionally, CD spectra of

mutant-type sequences were not affected in the presence of the G4 ligand, strongly indicating that wild-type sequences of *Dele* and *Cdc6* form G4 structures, and that corresponding selected mutants do not. These observations are in agreement with previous comparative analyses of mutant and wild type sequences.



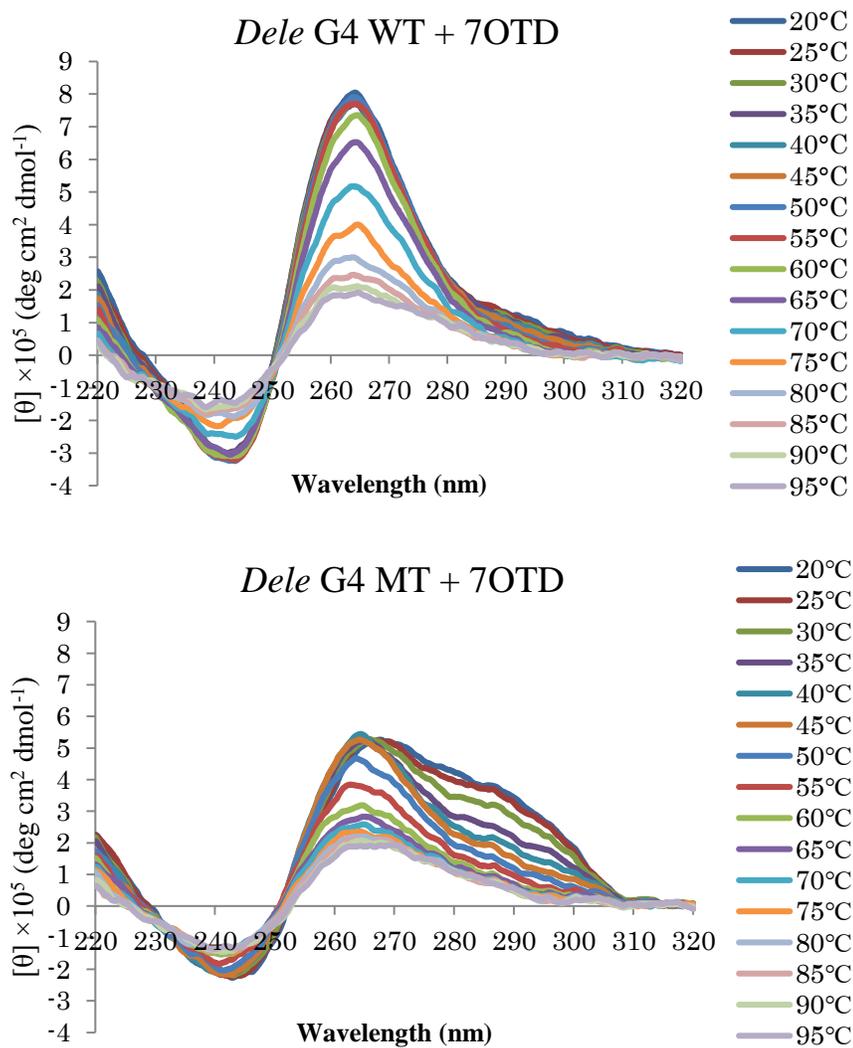
**Figure 2.11** CD spectra of wild-type and mutant *Dele* G4 structures

Wild-type (top) and mutant (bottom) G4 structures were analyzed in 5°C intervals from 20°C to 95°C in TK buffer containing 50-mM Tris-HCl and 100-mM KCl (pH 7.5).



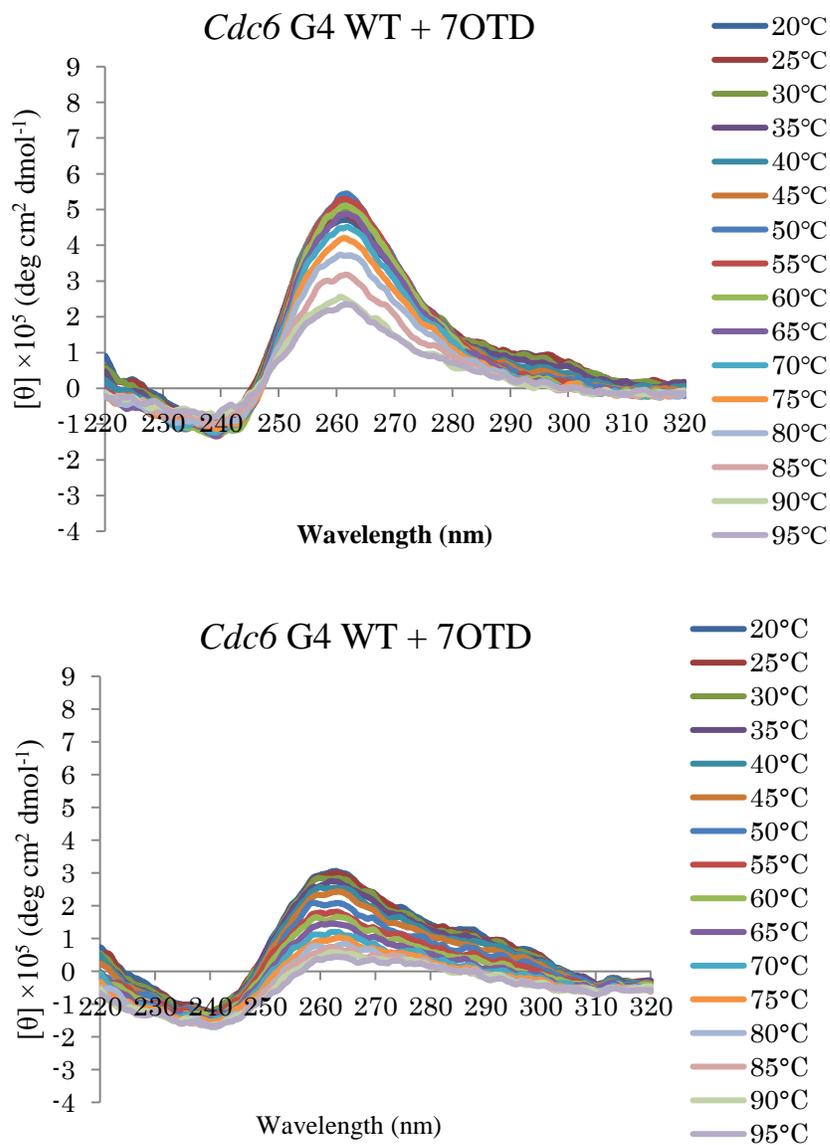
**Figure 2.12** CD spectra of wild-type and mutant *Cdc6* G4 structures

Wild-type (top) and mutant (bottom) structures were analyzed in 5°C intervals from 20°C to 95°C in TK buffer at pH 7.5.



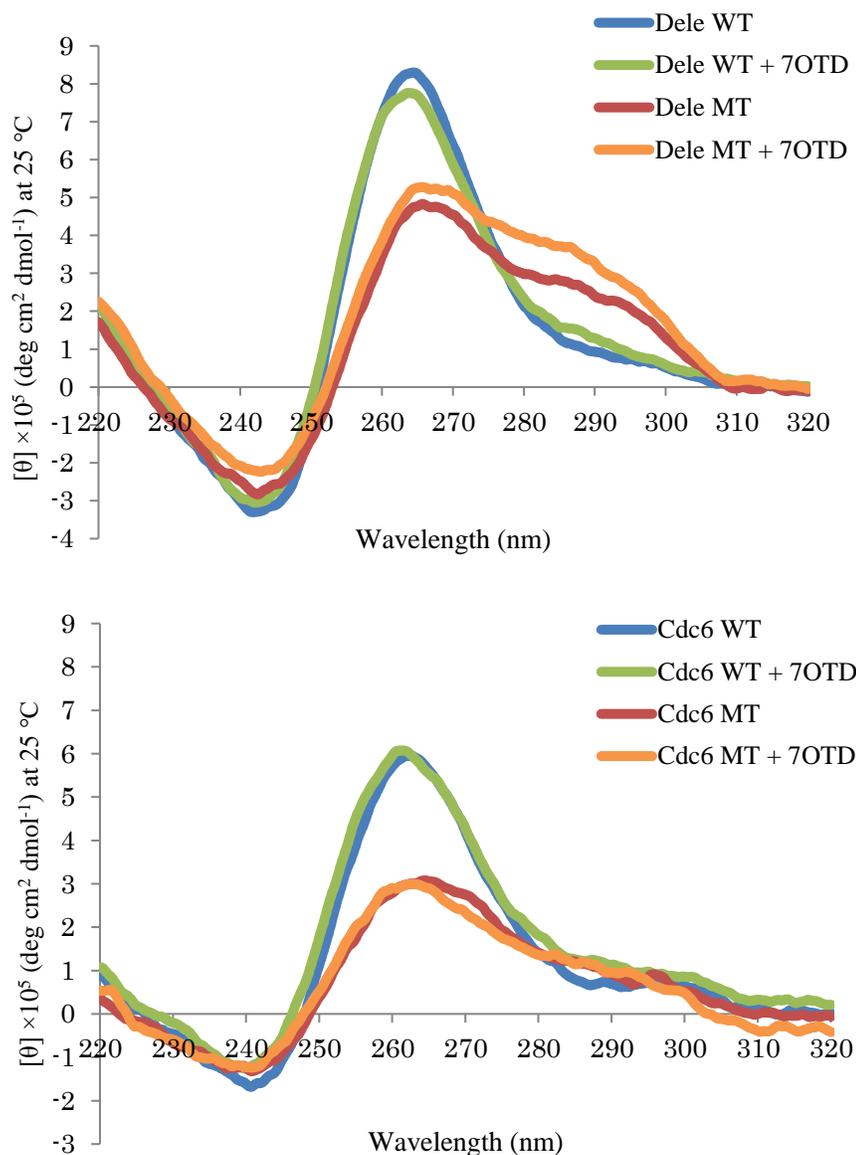
**Figure 2.13** CD spectra of wild-type and mutant *Dele G4* structures in the presence of 10- $\mu$ M L1H1-7OTD

Wild-type (top) and mutant (bottom) G4 structures were analyzed in 5°C intervals from 20°C to 95°C in TK buffer at pH 7.5.



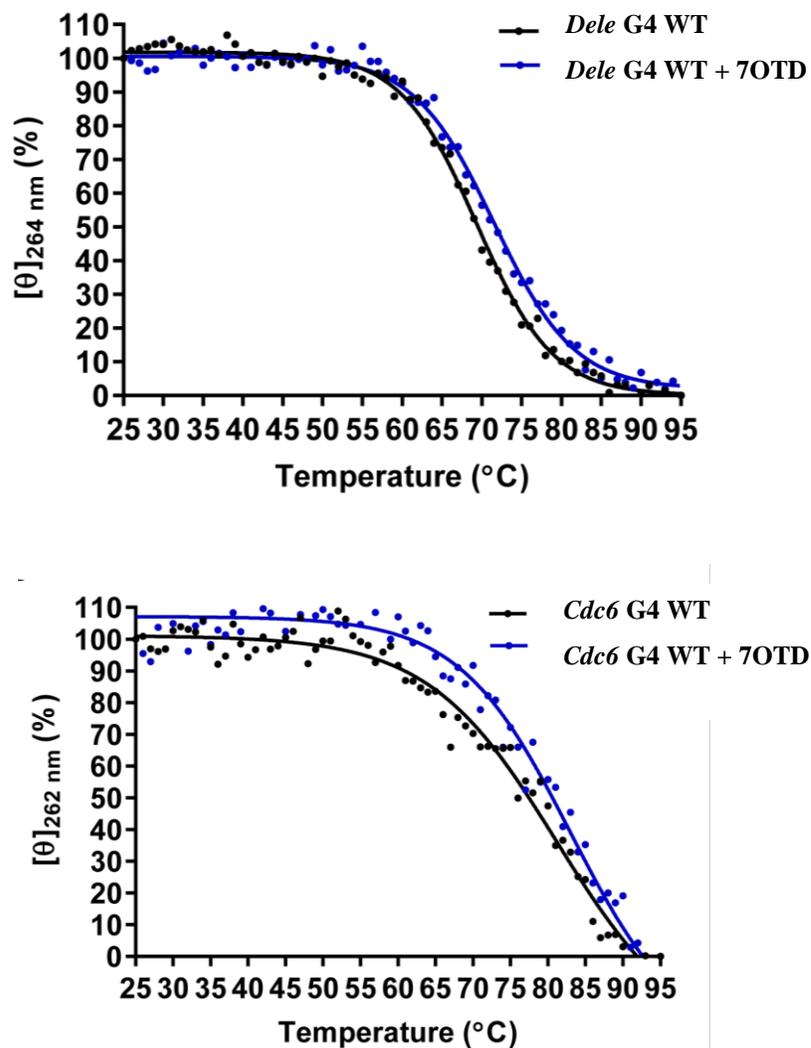
**Figure 2.14** CD spectra of the *Cdc6* G4 wild-type and *Cdc6* G4 mutant in the presence of 10  $\mu$ M L1H1-7OTD

Wild-type (top) and mutant (bottom) G4 structures were analyzed in 5°C intervals from 20°C to 95°C in TK buffer at pH 7.5.



**Figure 2.15** CD spectra of *Dele* G4 and *Cdc6* G4 DNAs.

Wild-type (blue), mutant (red), wild-type with L1H1-7OTD (green), and mutant with L1H1-7OTD (orange) were analyzed at 25°C in TK buffer at pH 7.5.



**Figure 2.16** CD melting curves of wild-type *Dele* G4 and *Cdc6* G4 DNAs at 264 and 262 nm, respectively

Wild-type DNAs were analyzed in the presence (blue) and the absence of L1H1-7OTD (black) in  $1^{\circ}\text{C}$  intervals from  $25^{\circ}\text{C}$  to  $95^{\circ}\text{C}$  using TK buffer at pH 7.5.

## 2.4 Summary and discussion

Herein, ten selected G4 DNAs were cloned into a luciferase reporter vector for reporter assays. These experiments showed that *Cdc6* and *Dele* G4 DNAs activate transcription, likely reflecting enhancer and promoter activities of these individual clones in the vector. However, binding of *Cdc6* and *Dele* G4 structures by L1H1-7OTD inhibited transcriptional activation, suggesting suppressive effects on enhancer–promoter interactions. In addition, previous reporter assays of CGI sequences showed that *Dele*-F G4 DNA possesses high transcriptional activity.

The present study demonstrated transcriptional activation by G4 structures and identified *Dele* and *Cdc6* as transcriptional regulators. Recently, the death ligand signal enhancer (DELE) was shown to bind death-associated protein 3 (DAP 3), which is induced by various stimuli to regulate cell apoptosis. Stable expression of DELE induces apoptosis, whereas knockdown of DELE rescued HeLa cells from apoptosis [91]. In addition, the cell-division-cycle 6 (CDC6) protein is essential for DNA replication, and downregulation of *Cdc6* expression leads to inhibition of cell growth and increased apoptosis [92]. Abnormal apoptosis is related to many diseases involving atrophy [93], such as Parkinson's disease [94], with excessive cell death leading to tissue and organ damage. In contrast, some oncogenic mutations disrupt apoptosis, leading to tumor progression [95] or metastasis. Therefore, further analysis of *Dele* and *Cdc6* G4 DNAs may contribute to the understanding of these apoptotic mechanisms.

## CHAPTER 3. Methylated G4 structures in transcription

### 3.1 Background and aims

DNA methylation is a well-known epigenetic modification that plays essential and global roles in the regulation of gene expression. Previous studies show that CpG methylation stabilizes G4 structures, with increased thermal stability of *Bcl-2* G4 structures [79]. Furthermore, CpG methylation stabilizes G4 structures of d(CGCG<sub>3</sub>GCG) oligonucleotides [96], *FMR1* repeats (d(CGG)<sub>5</sub>) [97], and *C9orf72* repeats (d(GGGGCC)<sub>8</sub>) [98], indicating that changes in the physical properties of DNA may affect transcriptional regulation. However, direct evidence of the effects of methylated G4 structures on transcription is lacking. Thus, transcriptional effects of stabilizing G4 structures through DNA methylation were investigated in this chapter.

## 3.2 Materials and methods

### 3.2.1 Preparation of methylated G4 DNA-containing plasmids for luciferase reporter assays

To analyze the transcriptional effects of methylated G4, two G4-forming sequences (Table 3.1) from mouse *Dele* and human *DELE* genes were cloned into the *Sfi*I site of the luciferase reporter vector pGL4.23. Subsequently, 2 µg aliquots of DNA plasmids were methylated using 8 U of CpG methyltransferase (NEB) and methylated and unmethylated preparations were then purified using phenol chloroform. Methylated and unmethylated DNA concentrations were determined spectrophotometrically (BioSpectrometer, Eppendorf), and to confirm methylation, 200 ng aliquots were treated with 5 U of the methylation-sensitive restriction enzyme *Hpa*II at 37°C for 1 h, and were then electrophoresed on 1% agarose gels in 1 × TAE buffer.

**Table 3.1** G4 DNA sequences for studies of the effects of methylation on transcription

CpG sites are highlighted.

Name	Sequences (5'–3')
Mouse <i>Dele</i> G4	GGGTGGGCTTAGATCTGGGAAGGGCGGG
Human <i>DELE</i> G4	GGGC <sup>CG</sup> GGTTAGAGCCAAGGAAGG <sup>CGGG</sup>

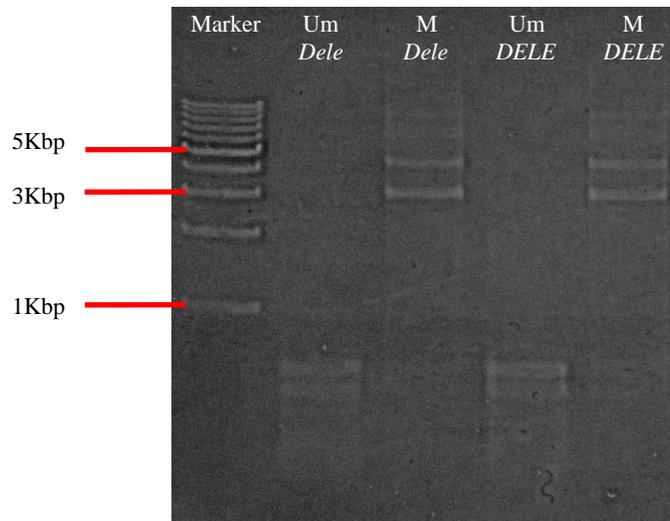
### **3.2.2 Cell culture and luciferase reporter assays**

NIH3T3 cells were cultured as described in chapter 2 (page 41), were seeded in 24 wells plate, and were transfected with 100 ng of firefly luciferase reporter vector and 100 ng of the Renilla luciferase control vector using Lipofectamine 3000 (Thermo Fisher Scientific, Waltham, MA, USA) according to the manufacturer's protocol. After 48 h, cells were lysed and luciferase activities were measured using the Dual-Luciferase Reporter Assay System (Promega Corporation, Madison, WI, USA) and a SPARK 10M microplate reader (TECAN, Männedorf, Switzerland). To determine firefly luciferase expression levels in cells, ratios of firefly to renilla luciferase activities were calculated, and were normalized to ratios of the pGL4.23 luciferase reporter vector. All reporter assays were performed in triplicate.

### 3.3 Results

#### 3.3.1 Confirmation of methylated DNA vectors

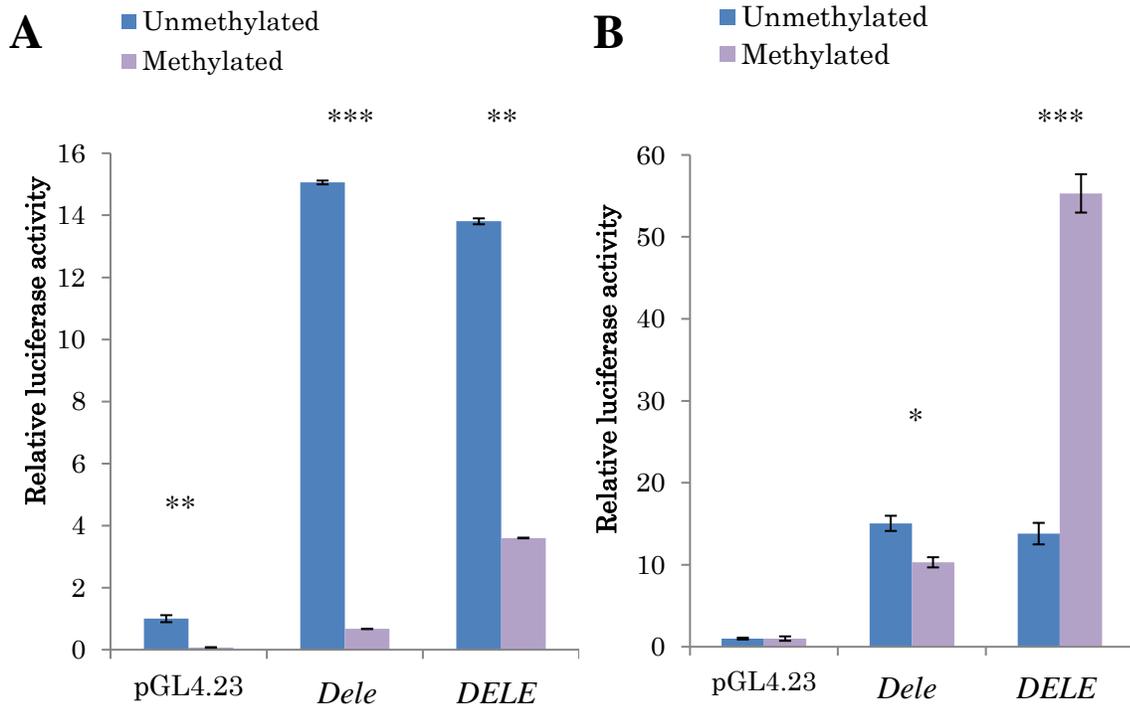
To confirm CpG methylation, the methylation-sensitive restriction enzyme *HpaII* was used to digest DNA plasmids. This restriction enzyme recognizes and cuts C/CGG sequences but is inactive on methylated ( $C^5$  mCGG) substrates. The plasmid pGL4.23 contains 24 CCGG sites, leading to the production of multiple fragments after digestion. In contrast, few fragments are produced after methylation of CGIs, as shown in Figure 3.1.



**Figure 3.1** Electrophoresis of *HpaII* digested template DNAs (red) was performed to confirm methylation

### 3.3.2 Effect of methylated G4 structures on transcription

G4 DNA sequences from mouse *Dele* and human *DELE* genes were cloned into luciferase reporter vectors, which were then transfected into NIH3T3 cells, and luciferase gene expression was measured after 48 h culture. Dual-luciferase assays were then performed and relative luciferase activities of methylated and unmethylated G4 vectors were measured and normalized to the activities of unmethylated controls. These experiments (Figure 3.2.A demonstrated an inhibition in expression of methylated G4 plasmids. However, because the methylated control vector had lower expression levels than the unmethylated control, expression levels of methylated and unmethylated G4 plasmids were normalized to respective controls (Figure 3.2.B). Under these conditions, methylation inhibited transcription of *Dele* G4 by 30% but increased that of *DELE* G4 to 290% of the control.



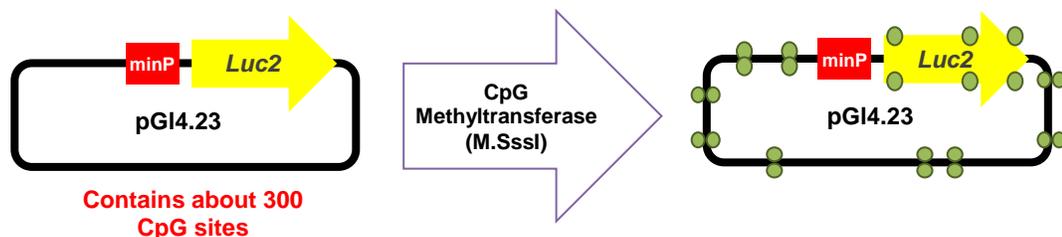
**Figure 3.2** Reporter assays showing the regulatory effects of CpG methylation on *Dele* and *DELE* G4 DNA expression; data were normalized to the unmethylated control (**A**) and to unmethylated and methylated controls (**B**).

Blue bars represent relative luciferase activities of unmethylated G4 DNAs, and violet bars represent relative luciferase activities of methylated G4 DNAs. Luciferase activities relative to that of pGL4.23 are presented as means  $\pm$  SD (n = 3). Differences between unmethylated and methylated samples were identified using t-tests; \*P < 0.01, \*\*P < 0.001, \*\*\*P < 0.0001

### 3.4 Summary and discussion

Methylated and unmethylated mouse and human *Dele* G4 structures inhibited transcription when normalized to unmethylated control vectors. However, when expression data for methylated G4 DNAs were normalized to the methylated control vector, transcription was inhibited by mouse *Dele* G4 but was enhanced by *DELE* G4 DNA.

The plasmid, pGL4.23, was used in this study but it contains about 300 CpG sites that can be methylated by CpG methyltransferase (Figure 3.3), and some of these may have affected the promoter on the vector, likely inhibiting expression of the control vector. Therefore, to specifically identify the effects of methylated G4 in the vector, synthetically methylated G4 sequences must be cloned onto the vector, or a CpG free vector must be used. The use of exclusively methylated G4 sequences may provide more accurate estimates of the effects of G4 methylation on transcription.



**Figure 3.3** The plasmid pGL4.23 contains numerous CpG sites, and all can be methylated.

## **CHAPTER 4. Utilizing G4 and i-motif forming sequences to detect DNA methylation**

### **4.1 Background and aims**

Several studies associate changes in methylation levels of specific genes with changes in expression that contribute to disease. Therefore, DNA methylation has been proposed as potential biomarker for cancer diagnosis. In this study, a new DNA methylation detection system was developed using G4 and i-motif forming sequences. This detection system does not require sodium bisulfite treatment or methylated DNA ligands, and can therefore be used to save time and money. G4 and i-motif structures that are stabilized by DNA methylation may arrest DNA polymerase activity during quantitative polymerase chain reaction (qPCR). Hence, G4 and i-motif forming sequences from *VEGF* and *RET* were used as templates in qPCR, and made comparisons based on PCR amplification efficiency.

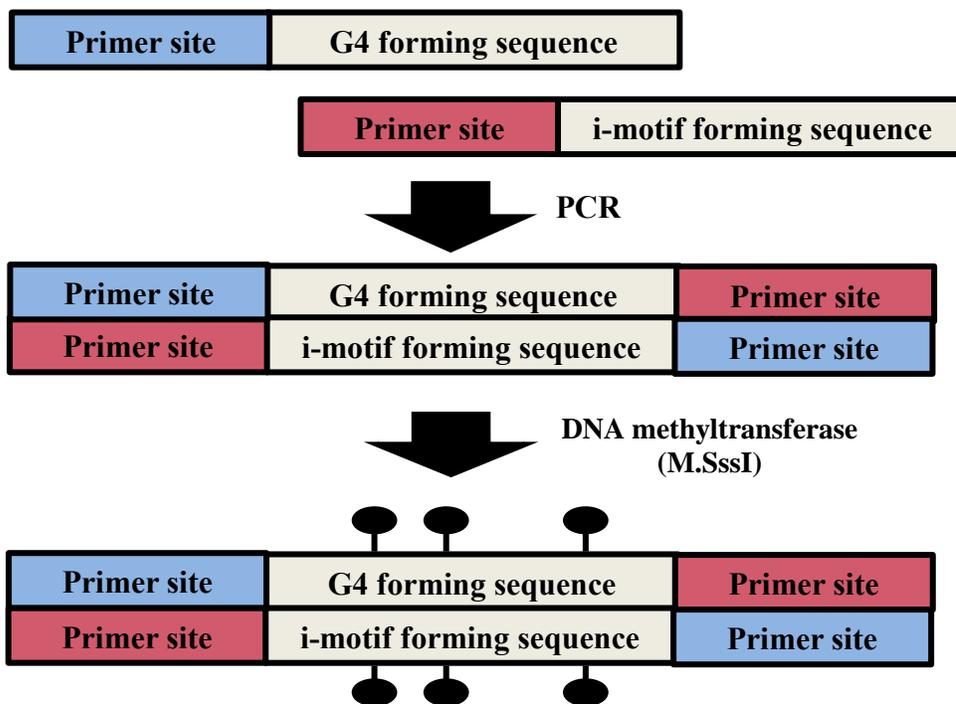
## 4.2 Materials and methods

### 4.2.1 Preparation of methylated G4 and i-motif DNA

To prepare template sequences, human genomic DNA was purified from HUVEC cells using DNeasy Blood & Tissue Kits (Qiagen). Subsequently, G4 regions in *VEGF*, *RET*, and *c-MYC* genes were amplified from genomic DNA (500 ng) template using PCR in 500  $\mu$ l solutions. G4 and i-motif forming sequences of *VEGF* [99] and *RET* [100] and *c-MYC* [101] were then amplified from the human genome using specific primers that were designed primers by Primer 3 (Figure 4.1, Table 4.1). In PCR reaction samples, Ex Taq HS (Takara) was added with buffer containing 25-mM TAPs, 2-mM MgCl<sub>2</sub>, 0.1-mM DTT, and 5% DMSO (pH 9.3). Thermocycling was initiated at 95°C for 5 min, followed by 35 cycles of 95°C for 30 s, 59°C for 30 s, and 72°C for 30 s. PCR products were then purified and 2  $\mu$ g aliquots were treated with 40 U of CpG methyltransferase (NEB). After methylation, DNA was purified using phenol chloroform and DNA concentrations were determined using a spectrophotometer (BioSpectrometer, Eppendorf). To confirm DNA methylation, 50 ng aliquots were treated with 3 U of the methylation-sensitive restriction enzyme *Hpa*II at 37°C for 1 h, and were then electrophoresed on 15% polyacrylamide gels in 1  $\times$  TBE buffer.

Mutant *VEGF* and *RET* G4 DNAs were prepared by performing overlap PCR using 3 fmol of the oligonucleotides *VEGF*\_MT\_F and *VEGF*\_MT\_R for *VEGF* mutant G4 DNA and *RET*\_MT\_F and *RET*\_MT\_R for *RET* mutant G4 DNA (Table 4.2). To amplify *VEGF* mutant G4 DNA, overlap PCR was performed using Ex Taq HS (Takara) with the supplied buffer. Thermocycling was performed with an initial step at 95°C for 5 min, followed by

20 cycles of 95°C for 30 s, 59°C for 30 s, and 72°C for 30 s. As for the *RET* mutant G4 DNA, overlap PCR was performed using Ex Taq HS (Takara) with buffer containing 25 mM tris(hydroxymethyl)methylamino]propanesulfonic acid (TAPs) (pH 9.3), 2 mM MgCl<sub>2</sub>, 0.1-mM DTT, and 5% DMSO. Thermocycling was performed with an initial step at 95°C for 5 min, followed by 25 cycles of 95°C for 30 s, 59°C for 30 s, and 72°C for 30 s. Finally, PCR products were methylated using the methods described above.



**Figure 4.1** Preparation of methylated target sequences

**Table 4.1** Primers for PCR amplification and overlap PCR

Name		Sequence (5'-3')
<i>VEGF</i>	Forward	GTCGAGCTTCCCCTTCATT
	Reverse	CGCTACCAGCCGACTTTT
<i>RET</i>	Forward	GGGACTGAAGCGAGGTAAGC
	Reverse	CGCACTGAGCTCCTACACG
<i>c-MYC</i>	Forward	CGGAGATTAGCGAGAGAGGA
	Reverse	TAGGCGCGCGTAGTTAATTC

**Table 4.2** Template DNAs for overlap PCR

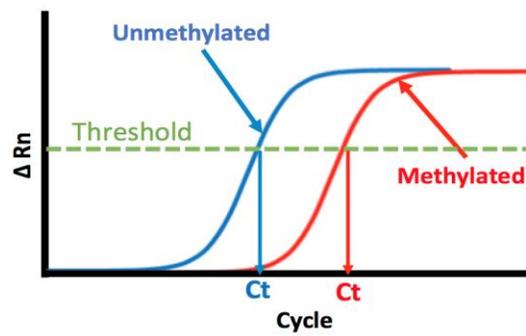
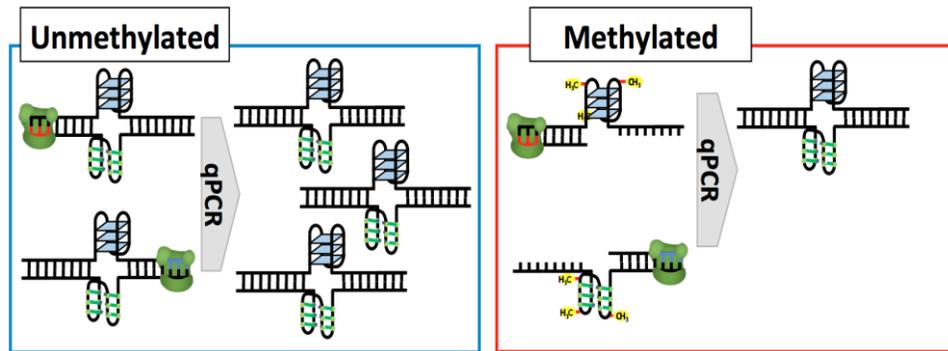
Name	Sequences (5'-3')
MT_ <i>VEGF</i> _F	GTCGAGCTTCCCCTTCATTGCGGGCGGGCTGCGGGCCAGGCTTCACTGAGCGTCCGC AGAGCCCGGGCCCGAGCCGCGTGTGGAAGGGCTGAGGCTCGCCT
MT_ <i>VEGF</i> _R	CGCTACCAGCCGACTTTTAAAAAAAAAAGGGGGGGCGCATGGCTCCGAAACGACG GGAAAACGAAAACGGAACGAACCGGGGGCGGGACAGGCGAGC
MT_ <i>RET</i> _F	GGGACTGAAGCGAGGTAAGCGCCGGCTGCGCCGGAGGAGCGGGTATTTGCGTTTCG TTGCGTTTGCAGGTCAGGGTGGGCCAGGCGGGCCGGAGGCGG
MT_ <i>VEGF</i> _R	CGCACTGAGCTCCTACACGCGCCGCCCCGCGCACCCCGCGCAGCCAGAGCAA GCACTGGAGCCCCGCCCTTCCCGCACCCACCCGCTCCGGCC

#### **4.2.2 Quantitative PCR analysis of methylated G4 and i-motif DNA**

Unmethylated and methylated DNA were mixed to prepare template DNA for qPCR at a methylation frequency of 0, 20, 40, 60, 80, and 100%. Quantitative PCR was then performed using a 7900 HT Fast Real-Time PCR System (ABI). Samples for qPCR were prepared in reaction volumes of 20  $\mu$ L containing primer pairs at 0.5  $\mu$ M, 4 pM of template DNA ( $1.0 \times 10^7$  copy), and SYBR Premix Ex Taq II Tli RNaseH Plus (Takara). Thermocycling was performed with an initial step at 95°C for 5 min, followed by 40 cycles of 95°C for 30 s, 59°C for 30 s, and 72°C for 30 s. To assess PCR amplification efficiency, changes in cycle threshold (Ct) values ( $\Delta$ Ct) relative to those for amplified unmethylated DNA template were analyzed. The DNA methylation detection system is presented in Figure 4.2.

#### **4.2.3 Detection limit of methylated G4 and i-motif DNA in qPCR analyses**

To investigate detection limits, unmethylated and methylated G4 and i-motif *VEGF* and *RET* DNA were diluted for use as qPCR templates. Samples containing  $1.0 \times 10^4$ ,  $1.0 \times 10^5$ ,  $1.0 \times 10^6$ ,  $1.0 \times 10^7$ ,  $1.0 \times 10^8$ , and  $1.0 \times 10^9$  copies of template DNAs were prepared and were analyzed using qPCR.



**Figure 4.2** Schematic of the present novel qPCR-based system for estimating DNA methylation

## 4.3 Results

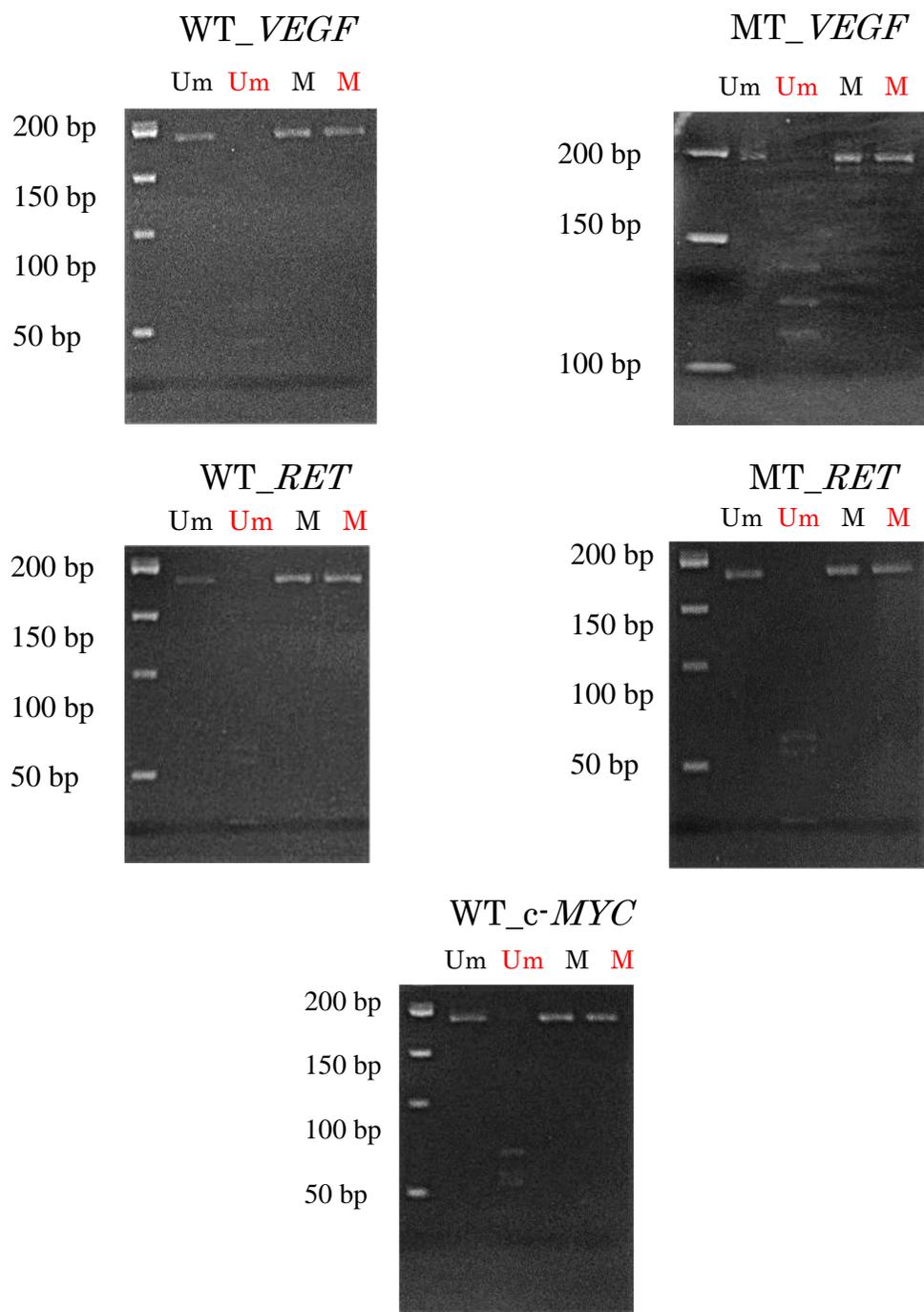
### 4.3.1 Preparation of template DNAs

*VEGF* and *RET* DNAs contain CpG sites on G4 and i-motif forming regions, whereas *c-MYC* DNAs do not and were used as a control. Guanine bases in mutant DNAs that are involved in the formation of G-tetrads were substituted with thymine bases using overlap PCR, except in CpG sequences. PCR products of mutant and wild-type *VEGF*, *RET*, and *c-MYC* DNAs were 192, 188, and 191-bp, respectively (Table 4.3). Methylation was performed using CpG methyl transferase, and was confirmed by digesting PCR products with the methylation-sensitive restriction enzyme *HpaII* followed by native polyacrylamide gel electrophoresis (PAGE; Figure 4.3). Finally, qPCR was performed using mixtures of unmethylated and methylated template DNAs containing 0%, 20%, 40%, 60%, 80%, and 100% methylated DNA.

**Table 4.3** Sequences of PCR products of *VEGF*, *RET*, and *c-MYC*

G4 and i-motif regions are boxed, guanine bases that form G-tetrads are shown in blue, mutation sites are shown in red, and CpG sequences are highlighted.

Name	Sequences (5'-3')
<i>VEGF</i>	GT <b>CG</b> AGCTTCCCCTTCATTG <b>CGG</b> CGGGCTG <b>CG</b> GGCCAGGCTTCACTGAG <b>CG</b> TC <b>CG</b> CAGAGCC <b>CG</b> GGCC <b>CG</b> AGC <b>CGCG</b> TGTGGAAGGGCTGAGGCT <b>CG</b> CCTGT CCC <b>CG</b> CCCC <b>CG</b> <b>GGGCGGGC</b> <b>CGGGGGCGGGG</b> TCC <b>CGG</b> <b>CG</b> GGG <b>CG</b> GAGCCAT <b>CG</b> CCCCCCCCCTTTTTTTTTTAAAAGT <b>CG</b> GCTGGTAG <b>CG</b>
MT_ <i>VEGF</i>	GT <b>CG</b> AGCTTCCCCTTCATTG <b>CGG</b> CGGGCTG <b>CG</b> GGCCAGGCTTCACTGAG <b>CG</b> TC <b>CG</b> CAGAGCC <b>CG</b> GGCC <b>CG</b> AGC <b>CGCG</b> TGTGGAAGGGCTGAGGCT <b>CG</b> CCTGT CCC <b>CG</b> CCCC <b>CG</b> <b>GTTTCGTTCCGTTTTCGTTT</b> TCCCGT <b>CG</b> TTT <b>CG</b> GAGCCAT <b>CG</b> CCCCCCCCCTTTTTTTTTTAAAAGT <b>CG</b> GCTGGTAG <b>CG</b>
<i>RET</i>	GGGACTGAAG <b>CG</b> AGGTAAG <b>CGC</b> CGGCTG <b>CGC</b> <b>CG</b> GAGGAG <b>CG</b> GGTA <b>GGGGCG</b> <b>GGGCGGGCGGGG</b> <b>CG</b> GTCCAGGGGTGGGCCAGG <b>CGGGGC</b> <b>CG</b> GAGG <b>CG</b> GGT GGGGTG <b>CG</b> GGAAGGGG <b>CG</b> GGGCTCCAGTGCTTGCTCTGGCTG <b>CGCG</b> GGGTG <b>CGGC</b> <b>CGGGGCGCG</b> <b>CGCG</b> TGTAGGAGCTCAGTG <b>CG</b>
MT_ <i>RET</i>	GGGACTGAAG <b>CG</b> AGGTAAG <b>CGC</b> CGGCTG <b>CGC</b> <b>CG</b> GAGGAG <b>CG</b> GGTA <b>TTTTCG</b> <b>TTTTCGTTGCGTTT</b> <b>CG</b> GTCCAGGGGTGGGCCAGG <b>CGGGGC</b> <b>CG</b> GAGG <b>CG</b> GGT GGGGTG <b>CG</b> GGAAGGGG <b>CG</b> GGGCTCCAGTGCTTGCTCTGGCTG <b>CGCG</b> GGGTG <b>CGGC</b> <b>CGGGGCGCG</b> <b>CGCG</b> TGTAGGAGCTCAGTG <b>CG</b>
<i>c-MYC</i>	<b>CG</b> GAGATTAG <b>CG</b> AGAGAGGATCTTTTTTCTTTTCCCC <b>CG</b> CCCTCTGCTT TGGGAACCC <b>CG</b> GGAGGG <b>CG</b> CTTATG <b>GGGAGGGTGGGGAGGG</b> TGGGGAAGGT GGGAGGAGACTCAGC <b>CG</b> GGCAGC <b>CG</b> AGCACTCTAGCTCTAGGATGTAAAC AGAGTAAGAGAGC <b>CG</b> CATGAATTA <b>ACTACGCGCG</b> CCTA

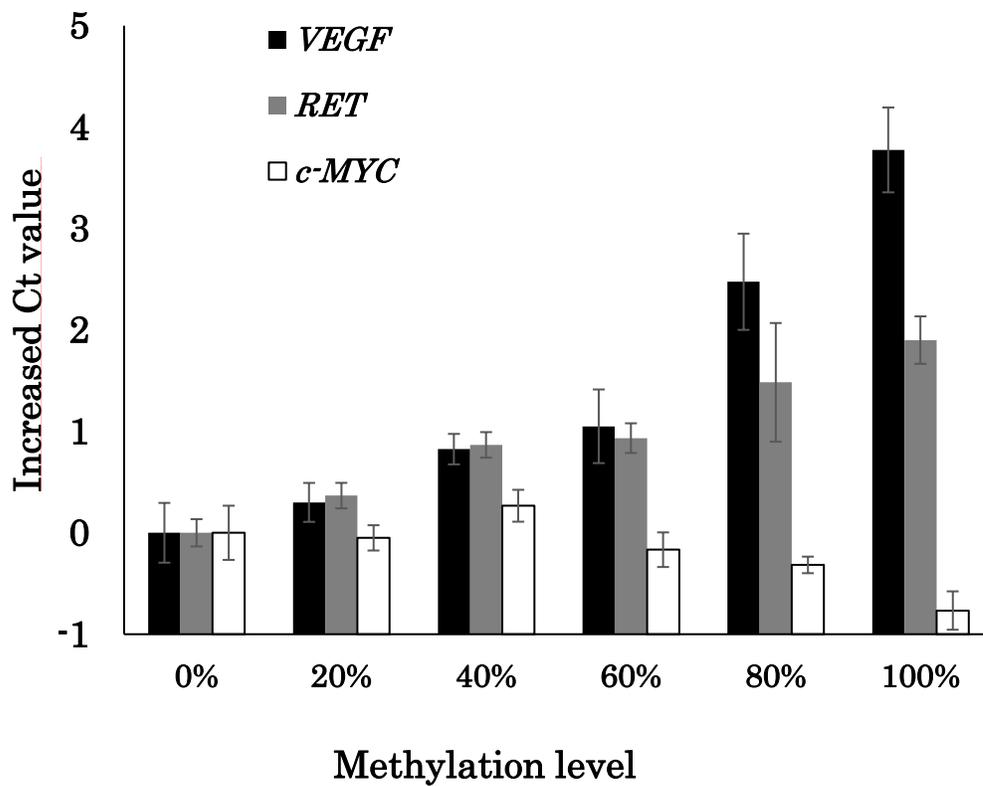


**Figure 4.3** Confirmation of methylation using electrophoresis analyses of *Hpa*II digested template DNAs (red)

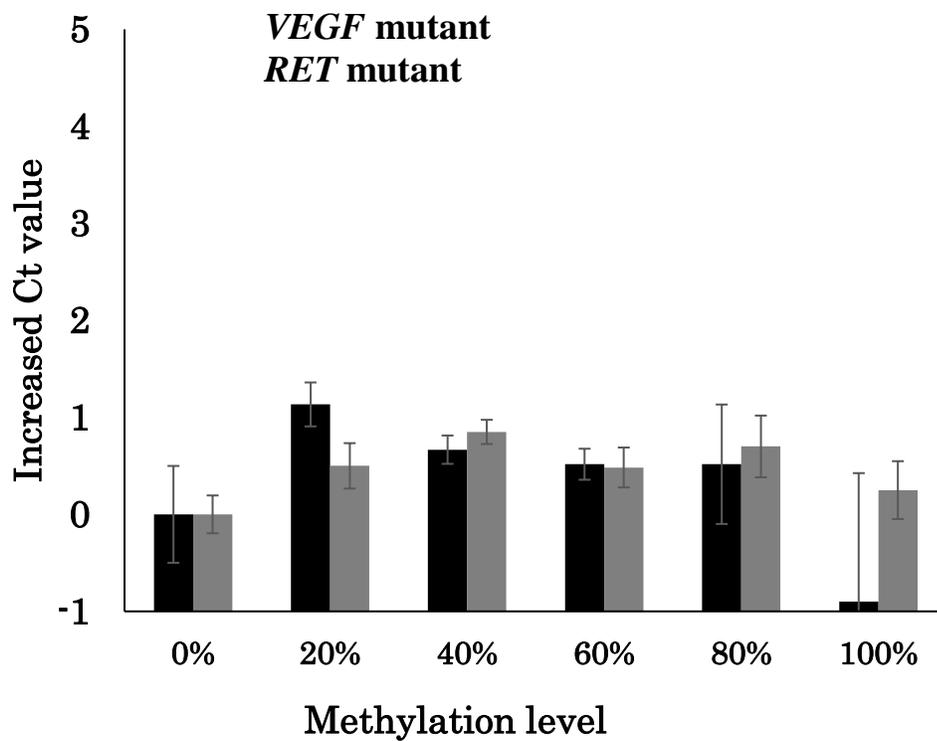
### 4.3.2 Methylated G4 and i-motif sequences affected qPCR amplification efficiency

In these experiments, increased  $\Delta C_t$  values indicate low amplification efficiency. Thus, methylation of G4 and i-motif *VEGF* and *RET* DNA templates led to decreased amplification efficiency. (Figure 4.4). However, qPCR analyses of *c-MYC* DNAs, which lack CpG sequences in G4 and i-motif forming sequences, showed no correlation between  $\Delta C_t$  values and methylation levels (Figure 4.4). In contrast, PCR amplification efficiency was negatively correlated with DNA methylation levels of *VEGF* and *RET* DNAs, which contain CpG sites on G4 and i-motif forming sequences.

In subsequent experiments, we determined amplification efficiency of mutant *VEGF* and *RET* DNAs that do not form G4 and i-motif structures, and observed no correlation between methylation levels and  $\Delta C_t$  values (Figure 4.5). Although mutant *VEGF* and *RET* DNAs contained GC-rich sequences with GC contents of 66% and 74%, respectively, PCR efficiency of methylated templates did not differ from that of unmethylated templates. These data indicate that methylated GC-rich sequences containing no G4 and i-motif forming sequences do not effect PCR efficiency, whereas methylated G4 and i-motif forming sequences reduce PCR amplification efficiency. Hence, methylation levels of G4 and i-motif forming sequences in template DNAs were detectable using qPCR.



**Figure 4.4** Quantitative PCR analysis of methylated G4 and i-motif-forming sequences  
 Quantitative PCR analyses of wild-type *VEGF* (black), *RET* (gray), and *c-MYC* (white) G4 and i-motif DNAs; increased Ct values of unmethylated DNA are presented relative to those of respective methylated DNAs (mean  $\pm$  SD, n = 4).



**Figure 4.5** Quantitative PCR analysis of methylated mutant G4 and i-motif-forming sequences

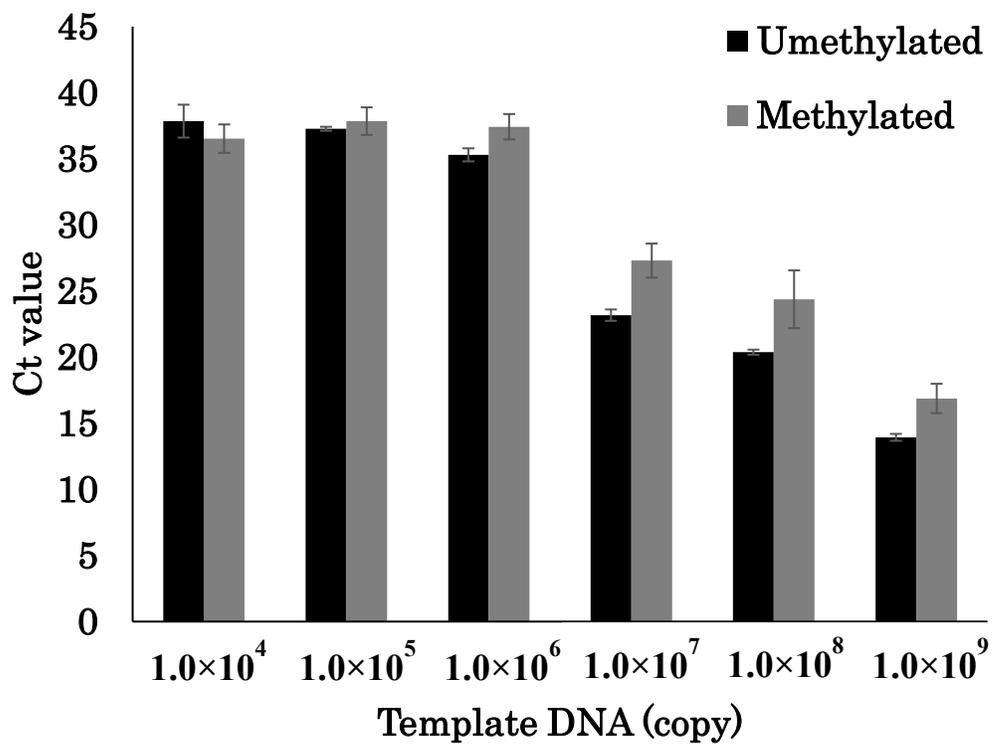
Quantitative PCR analyses of mutant *VEGF* (black), *RET* (gray), and *c-MYC* (white) G4 and i-motif DNAs; increased Ct values are presented relative to those of respective unmethylated DNAs (mean  $\pm$  SD, n = 4).

Ct values of 0% methylated *RET* DNA differed significantly from those of 20% methylated *RET* DNA ( $P < 0.05$ ). However,  $\Delta$ Ct values were insufficient to detect hypomethylation. In a previous study, DNA methylation levels of various regions differed by  $> 50\%$  between normal and cancer cells [68], indicating that the sensitivity of the present system for detecting DNA methylation is sufficient for cancer diagnosis.

### 4.3.3 Detection limit analysis

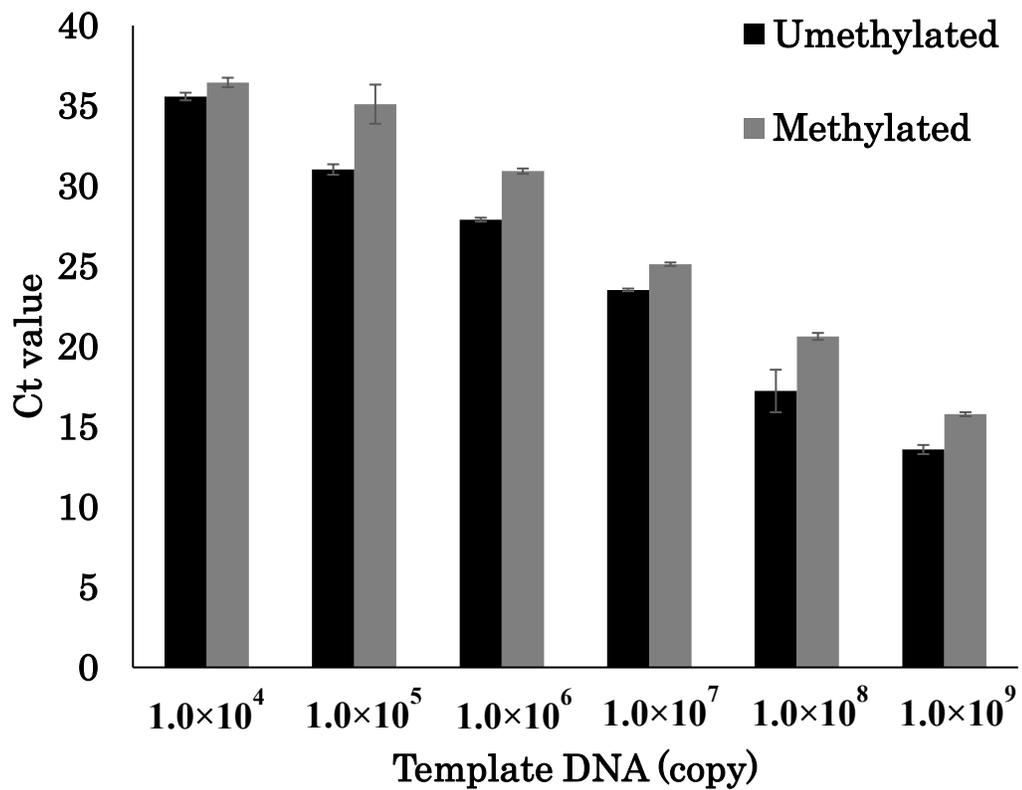
In the present study, methylated and unmethylated DNA templates had differing amplification efficiencies at  $1.0 \times 10^6$  to  $1.0 \times 10^9$  copies of *VEGF* DNA (Figure 4.6), and at  $1.0 \times 10^5$  to  $1.0 \times 10^9$  copies of *RET* DNA (Figure 4.7). Accordingly, respective detection limits were  $1.0 \times 10^6$  and  $1.0 \times 10^5$  copies. The present system for estimating DNA methylation levels indicated that  $\Delta C_t$  values depend on increased stability of G4 and i-motifs by DNA methylation, whereas detection limits are dependent on the amplification efficiency of unmethylated DNA. In accordance, unmethylated *VEGF* DNA was not amplified when present at  $1.0 \times 10^5$ , although the same number of unmethylated *RET* DNA copies were amplified, indicating a lower detection limit for methylated *RET* DNA.

Because  $C_t$  values are sensitive to template DNA concentrations, prior knowledge of genomic DNA concentrations is necessary to normalize estimates of methylation frequencies of *VEGF* and *RET* regions on genomic DNA. However, amplification efficiency of *c-MYC* was not dependent on DNA methylation, allowing determination of genomic DNA concentrations and elimination of false positives that are caused by polymerase inhibitors.



**Figure 4.6** Relationships between Ct values and template DNA copy numbers of methylated *VEGF* G4 and i-motif-forming sequences

Quantitative PCR were performed with  $1.0 \times 10^4$  to  $1.0 \times 10^9$  copies of *VEGF* G4 and i-motif DNA (mean  $\pm$  SD, n = 4)



**Figure 4.7** Relationships between Ct values and template DNA copy numbers of methylated *RET* G4 and i-motif-forming sequences

Quantitative PCR were performed using  $1.0 \times 10^4$  to  $1.0 \times 10^9$  copies of *RET* G4 and i-motif DNA (mean  $\pm$  SD, n = 4)

#### **4.4 Summary and discussion**

PCR amplification efficiencies decreased with increased DNA methylation levels of *VEGF* and *RET* DNAs, which contain CpG sites on G4 and i-motif forming sequences. Hence, the present data demonstrate that methylated G4 structures can be used to detect DNA methylation levels. Previous genome-wide DNA methylation analyses show that DNA methylation usually occurs in CGIs [102], which are highly concentrated in promoter regions. In addition, G4-forming sequences were previously identified in CGIs using DNA microarray analyses with a G4 ligand [58]. Because the human genome reportedly contains 716,310 G4 forming sequences [103], methylated G4 and i-motif structures may be present in several promoter regions, and could be diagnostic for various diseases. In addition, determinations of PCR amplification efficiencies can be used to detect DNA methylation on various promoter regions.

## CHAPTER 5. Conclusions and prospects

*Dele* and *Cdc6* G4 DNAs have high transcriptional activity when cloned individually on reporter vectors, and may play roles as promoters and enhancers of transcription. Herein, the formation of G4 in wild-type sequences was verified using CD spectroscopy, and transcriptional activation of *Dele-F* and *Cdc6* G4 DNAs was inhibited by 35%–37% in the presence of the telomestatin derivative L1H1-7OTD. CD spectral analyses also demonstrated that binding of L1H1-7OTD stabilizes G4 structures, and may modulate interactions between G4 structures and transcription factors. Although *Dele-F* G4 DNA was transcribed even within CpG islands, transcriptional activities of CGI sequences from *Cdc6* G4 DNA were different from those of individual clones. Thus, *Cdc6* G4 may have dual functional roles in transcriptional regulation, suggesting dual regulatory roles of G4 structures. Taken together, these data suggest that *Dele* and *Cdc6* G4 structures form under physiological conditions and regulate transcription.

Methylation of mouse *Dele* and human *DELE* G4 DNAs inhibited transcription when normalized to unmethylated control vectors, although methylation of control vectors was also inhibitory. These observations may reflect high numbers of CpG sites in the vector, resulting in methylation of the whole vector and the G4 area. After normalizing methylated G4 DNAs to methylated control vectors, transcriptional activation of the methylated mouse gene *Dele* was inhibited by 30%, whereas that of human *DELE* G4 DNA was controversially increased by 290%. However, because DNA methylation is widely considered inhibitory of transcription, these observations require confirmation using a specifically methylated G4 region and an unmethylated vector.

Herein, we assessed the effects of DNA methylation on PCR amplification efficiency using G4/i-motif forming sequences from the genes *VEGF*, *RET*, and *c-MYC*. In these experiments,  $\Delta C_t$  values increased with methylation levels, demonstrating a negative correlation between qPCR amplification efficiency and DNA methylation levels in *VEGF* and *RET* DNAs. This correlation warrants further consideration of qPCR amplification efficiency as a tool for determining DNA methylation levels. In the present experiments, detection limits for *VEGF* and *RET* DNAs were  $1.0 \times 10^6$  and  $1.0 \times 10^5$  copies, respectively, in 20  $\mu$ L reaction volumes. However, this method could be improved by optimization at the nL scale, such as in the MicroTAS-based PCR system [104], which would allow single step PCR determinations of DNA methylation levels with fewer copies. In addition, DNA methylation was previously determined according to PCR amplification efficiency using a 3'-mismatched primer with a mutant DNA polymerase [105]. These studies showed that the 3'-mismatched primer was more efficiently extended from methylated than unmethylated cytosine residues, suggesting that PCR amplification efficiency increases with methylation. In contrast, PCR amplification efficiency was negatively correlated with template DNA methylation in our hands. Nonetheless, future studies are warranted using the present DNA methylation detection system with 3'-mismatched primers and mutant DNA polymerase, and will likely led to the development of an accurate system for determining methylation levels.

## BIBLIOGRAPHY

1. Wells RD. (1988) Unusual DNA structures. *J Biol Chem.* 263:1095–1098
2. Watson JD, Crick FH. (1953) Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. *Nature.* 171:737–738
3. Honig B, Rohs R. (2011) Biophysics: Flipping Watson and Crick. *Nature.* 470:472–473
4. Qin Y, Hurley LH. (2008) Structures, folding patterns, and functions of intramolecular DNA G-quadruplexes found in eukaryotic promoter regions. *Biochimie.* 90:1149–1171
5. Hardin CC, Watson T, Corregan M, Bailey C. (1992) Cation-dependent transition between the quadruplex and Watson-Crick hairpin forms of d(CGCG3GCG). *Biochemistry.* 31:833–841
6. Phan AT, Mergny JL. (2002) Human telomeric DNA: G-quadruplex, i-motif and Watson-Crick double helix. *Nucleic Acids Res.* 30:4618–4625
7. Simone R, Fratta P, Neidle S, Parkinson GN, Isaacs AM. (2015) G-quadruplexes: Emerging roles in neurodegenerative diseases and the non-coding transcriptome. *FEBS Lett.* 589:1653–1668
8. Gellert M, Lipsett MN, Davies DR. (1962) Helix formation by guanylic acid. *Proc Natl Acad Sci USA.* 48:2013–2018

9. Parkinson GN (2006). Fundamentals of quadruplex structures. In: Neidel S, Balasubramanian S (eds). Quadruplex nucleic acids. *Royal Society of Chemistry, Cambridge*.
10. Williamson JR, Raghuraman MK, Cech TR. (1989) Monovalent cation-induced structure of telomeric DNA: the G-quartet model. *Cell*. 59:871–880
11. Parkinson GN, Lee MP, Neidle S. (2002) Crystal structure of parallel quadruplexes from human telomeric DNA. *Nature*. 417:876–880
12. Duquette ML, Handa P, Vincent JA, Taylor AF, Maizels N. (2004) Intracellular transcription of G-rich DNAs induces formation of G-loops, novel structures containing G4 DNA. *Genes Dev*. 18:1618–1629
13. Maizels N, Gray LT. (2013) The G4 genome. *PLoS Genet*. 9:e1003468
14. Agrawal P, Hatzakis E, Guo K, Carver M, Yang D. (2013) Solution structure of the major G-quadruplex formed in the human VEGF promoter in K<sup>+</sup>: insights into loop interactions of the parallel G-quadruplexes. *Nucleic Acids Res*. 41:10584–10592
15. Karsisiotis AI, Hessari NM, Novellino E, Spada GP, Randazzo A, Webba da Silva M. (2011) Topological characterization of nucleic acid G-quadruplexes by UV absorption and circular dichroism. *Angew Chem Int Ed Engl*. 50:10645–10648
16. Marchand A, Gabelica V. (2016) Folding and misfolding pathways of G-quadruplex DNA. *Nucleic Acids Research*. 44:10999–11012
17. Dai J, Carver M, Punchihewa C, Jones RA, Yang D. (2007) Structure of the Hybrid-2 type intramolecular human telomeric G-quadruplex in K<sup>+</sup> solution: insights into

- structure polymorphism of the human telomeric sequence. *Nucleic Acids Research*. 35:4927–4940
18. Burge S, Parkinson GN, Hazel P, Todd AK, Neidle S. (2006) Quadruplex DNA: sequence, topology and structure. *Nucleic Acids Research*. 34: 5402–5415
  19. Patel DJ, Phan AT, Kuryavi V. (2007) Human telomere, oncogenic promoter and 5'-UTR G-quadruplexes: diverse higher order DNA and RNA targets for cancer therapeutics. *Nucleic Acids Research*. 35:7429–7455
  20. Balasubramanian S, Hurley LH, Neidle S. (2011) Targeting G-quadruplexes in gene promoters: a novel anticancer strategy?. *Nat Rev Drug Discov*. 10:261–275
  21. D'Antonio L, Bagga PS. (2004) Computational Methods for Predicting Intramolecular G-Quadruplexes in Nucleotide Sequences. *IEEE Computational Systems Bioinformatics Conference*. 590–591
  22. Hershman SG, Chen O, Lee JY, Kozak ML, Yue P, Wang L, Johnson FB. (2008) Genomic distribution and functional analyses of potential G-quadruplex-forming sequences in *Saccharomyces cerevisiae*. *Nucleic Acids Research*. 36:144–156
  23. Cao K, Ryvkin P, Johnson FB. (2012) Computational Detection and Analysis of Sequences with Duplex-Derived Interstrand G-quadruplex Forming Potential. *Methods*. 57: 3–10
  24. Kudlicki AS. (2016) G-Quadruplexes Involving Both Strands of Genomic DNA Are Highly Abundant and Colocalize with Functional Sites in the Human Genome. *PLoS ONE*. 11:e0146174

25. Harrington C, Lan Y, Akman SA. (1997) The identification and characterization of a G4-DNA resolvase activity. *J Biol Chem.* 272:24631–24636
26. Zhang Z, He X, Yuan G. (2011) Regulation of the equilibrium between G-quadruplex and duplex DNA in promoter of human c-myc oncogene by a pyrene derivative. *Int J Biol Macromol.* 49:1173–1176
27. Huppert JL, Balasubramania S. (2005) Prevalence of quadruplexes in the human genome. *Nucleic Acids Research.* 33, 2908–2916
28. Siddiqui-Jain A, Grand CL, Bearss DJ, Hurley LH. (2002) Direct evidence for a G-quadruplex in a promoter region and its targeting with a small molecule to repress c-MYC transcription. *Proc Natl Acad Sci USA.* 99:11593–11598
29. Rankin S, Reszka AP, Huppert J, Zloh M, Parkinson GN, Todd AK, Ladame S, Balasubramanian S, Neidle S. (2005) Putative DNA Quadruplex Formation within the Human c-kit Oncogene. *J Am Chem Soc.* 127: 10584–10589
30. Dai J, Dexheimer TS, Chen D, Carver M, Ambrus A, Jones RA, Yang D. (2006) An Intramolecular G-Quadruplex Structure with Mixed Parallel/ Antiparallel G-strands Formed in the Human BCL-2 Promoter Region in Solution. *J Am Chem Soc.* 128: 1096–1098
31. Guo K, Pourpak A, Beetz-Rogers K, Gokhale V, Sun D, Hurley<sup>[11]</sup><sub>[SEP]</sub>LH. (2007) Formation of pseudo-symmetrical G-quadruplex and i-motif structures in the proximal promoter region of the *RET* oncogene. *J Am Chem Soc.* 129:10220–10228

32. Cogoi S, Shchekotikhin AE, Xodo LE. (2014) *HRAS* is silenced by two neighboring G-quadruplexes and activated by MAZ, a zinc-finger transcription factor with DNA unfolding property. *Nucleic Acids Research*. 42:8379–8388
33. Membrino A, Cogoi S, Pedersen EB, Xodo LE (2011) G4-DNA Formation in the *HRAS* Promoter and Rational Design of Decoy Oligonucleotides for Cancer Therapy. *PLoS ONE* 6: e24421
34. Brito H, Martins AC, Lavrado J, Mendes E, Francisco AP, Santos SA, Ohnmacht SA, Kim NS, Rodrigues CM, Moreira R, Neidle S, Borralho PM, Paulo A. (2015) Targeting *KRAS* Oncogene in Colon Cancer Cells with 7-Carboxylate Indolo[3,2-b]quinoline Tri-Alkylamine Derivatives. *PLoS ONE*. 10:e0126891
35. Huppert JL, Balasubramanian S. (2007) G-quadruplexes in promoters throughout the human genome. *Nucleic Acids Res*. 35:406–413
36. Zhao Y, Du Z, Li N. (2007) Extensive selection for the enrichment of G4 DNA motifs in transcriptional regulatory regions of warm blooded animals. *FEBS Lett*. 581:1951–1956
37. Eddy J, Maizels N. (2008) Conserved elements with potential to form polymorphic G-quadruplex structures in the first intron of human genes. *Nucleic Acids Res*. 36:1321–1333
38. Chariker JH, Miller DM, Rouchka EC. (2016) Computational Analysis of G-Quadruplex Forming Sequences across Chromosomes Reveals High Density Patterns Near the Terminal Ends. *PLoS ONE* 11: e0165101

39. Bochman ML, Paeschke K, Zakian VA. (2012) DNA secondary structures: stability and function of G-quadruplex structures. *Nat Rev Genet.* 13:770–780
40. Wang Q, Liu J, Chen Z, Zheng, Chen C, Hao Y, Tan Z. (2011) G-Quadruplex Formation at the 3' End of Telomere DNA Inhibits Its Extension by Telomerase, Polymerase and Unwinding by Helicase. *Nucleic Acids Research.* 39:6229–6237
41. Verma A, Halder K, Halder R, Yadav V, Rawal P, Thakur RK, Mohd F, Sharma A, Chowdhury S. (2008) Genome-wide computational and expression analyses reveal G-quadruplex DNA motifs as conserved cis-regulatory elements in human and related species. *J Med Chem.* 51:5641–5649
42. Brooks TA, Kendrick S, Hurley LH. (2010) Making sense of G-quadruplex and i-motif functions in oncogene promoters. *FEBS J.* 277:3459–3469
43. Borgognone M, Armas P, Calcaterra NB. (2010) Cellular nucleic-acid-binding protein, a transcriptional enhancer of c-Myc, promotes the formation of parallel G-quadruplexes. *Biochem J.* 428:491–498
44. Brooks TA, Hurley LH. (2010) Targeting MYC Expression through G-Quadruplexes. *Genes & Cancer.* 1:641–649
45. Seenisamy J, Rezler EM, Powell TJ, Tye D, Gokhale V, Joshi CS, Siddiqui-Jain A, Hurley LH. (2004) The dynamic character of the G-quadruplex element in the c-MYC promoter and modification by TMPyP4. *J Am Chem Soc.* 126:8702–8709
46. Gonzalez V, Guo K, Hurley L, Sun D. (2009) Identification and characterization of nucleolin as a c-myc G-quadruplex-binding protein. *J Biol Chem.* 284:23622–23635

47. Wang XD, Ou TM, Lu YJ, Li Z, Xu Z, Xi C, Tan JH, Huang SL, An LK, Li D, Gu LQ, Huang ZS. (2010) Turning off transcription of the *bcl-2* gene by stabilizing the *bcl-2* promoter quadruplex with quindoline derivatives. *J Med Chem Soc.* 53:4390–4398
48. Dexheimer TS, Sun D, Hurley LH. (2006) Deconvoluting the structural and drug-recognition complexity of the G-quadruplex-forming region upstream of the *bcl-2* P1 promoter. *J Am Chem Soc.* 128:5404–5415
49. Wu Y, Zan LP, Wang XD, Lu YJ, Ou TM, Lin J, Huang ZS, Gu LQ. (2014) Stabilization of *VEGF* G-quadruplex and inhibition of angiogenesis by quindoline derivatives. *Biochim Biophys Acta.* 1840:2970–2977
50. Sun H, Xiang J, Shi Y, Yang Q, Guan A, Li Q, Yu L, Shang Q, Zhang H, Tang Y, Xu G. (2014) A newly identified G-quadruplex as a potential target regulating Bcl-2 expression. *Biochim Biophys Acta.* 1840:3052–3057
51. Dhakal S, Yu Z, Konik R, Cui Y, Koirala D, Mao H. (2012) G-quadruplex and i-motif are mutually exclusive in ILPR double-stranded DNA. *Biophys J.* 102:2575–2584
52. Connor AC, Frederick KA, Morgan EJ, McGown LB. (2006) Insulin capture by an insulin-linked polymorphic region G-quadruplex DNA oligonucleotide. *J Am Chem Soc.* 128:4986–4991
53. Verma A, Yadav VK, Basundra R, Kumar A, Chowdhury S. (2009) Evidence of genome-wide G4 DNA-mediated gene expression in human cancer cells. *Nucleic Acids Res.* 37:4194–4204

54. Catasti P, Chen X, Moyzis RK, Bradbury EM, Gupta G. (1996) Structure-function correlations of the insulin-linked polymorphic region. *J Mol Biol.* 264:534–545
55. Timmer CM, Michmerhuizen NL, Witte AB, Van Winkle M, Zhou D, Sinniah K. (2014) An isothermal titration and differential scanning calorimetry study of the G-quadruplex DNA-insulin interaction. *J Phys Chem B.* 118:1784–1790
56. Palumbo SL, Memmott RM, Uribe DJ, Krotova-Khan Y, Hurley LH, Ebbinghaus SW. (2008) A novel G-quadruplex-forming GGA repeat region in the c-myc promoter is a critical regulator of promoter activity. *Nucleic Acids Res.* 36:1755–1769
57. Balasubramanian S, Neidle S. (2009) G-quadruplex nucleic acids as therapeutic targets. *Curr Opin Chem Biol.* 13:345–353
58. Iida K, Nakamura, T, Yoshida W, Tera M, Nakabayashi K, Hata K, Ikebukuro K, Nagasawa K. (2013) Fluorescent-ligand-mediated screening of G-quadruplex structures using a DNA microarray. *Angew Chem Int Ed Engl.* 52:12052–12055.
59. A Mosquera Orgueira. (2015) Hidden among the crowd: differential DNA methylation-expression correlations in cancer occur at important oncogenic pathways. *Front Genet.* 6:163
60. Pauly GT, Moschel RC. (2001) Mutagenesis by O<sup>6</sup>-Methyl-, O<sup>6</sup>-Ethyl-, and O<sup>6</sup>-Benzylguanine and O<sup>4</sup>-Methylthymine in Human Cells: Effects of O<sup>6</sup>-Alkylguanine-DNA Alkyltransferase and Mismatch Repair. *Chem Res Toxicol.* 14:894–900
61. Sukumar S, Notario V, Martin-Zanca D, Barbacid M. (1983) Induction of mammary carcinomas in rats by nitroso-methylurea involves malignant activation of H-ras-1 locus by single point mutations. *Nature.* 306:658–661

62. Zhao A, Zhao C, Tateishi-Karimata H, Ren J, Sugimoto N, Qu X. (2016) Incorporation of O<sup>6</sup>-methylguanine restricts the conformational conversion of the human telomere G-quadruplex under molecular crowding conditions. *Chem. Commun.* 52:1903
63. Saxonov S, Berg P, Brutlag DL. (2006) A genome-wide analysis of CpG dinucleotides in the human genome distinguishes two distinct classes of promoters. *Proc Natl Acad Sci U S A.* 103:1412–1417
64. Jones MJ, Goodman SJ, Kobor MS. (2015) DNA methylation and healthy human aging. *Aging Cell.* 14:924–932
65. Daura-Oller E, Cabre M, Montero MA, Paternain JL, Romeu A. (2009) Specific gene hypomethylation and cancer: new insights into coding region feature trends. *Bioinformatics.* 3:340–343
66. Li Y, Tollefsbol TO. (2011) DNA methylation detection: bisulfite genomic sequencing analysis. *Methods Mol Biol.* 791:11–21
67. Tan AC, Jimeno A, Lin SH, Wheelhouse J, Chan F, Solomon A, Rajeshkumar NV, Rubio-Viqueira B, Hídalgo M. (2009) Characterizing DNA methylation patterns in pancreatic cancer genome. *Mol Oncol.* 3:425–438
68. Mikeska T, Craig JM. (2014) DNA methylation biomarkers: cancer and beyond. *Genes.* 5:821–864
69. Gyparaki MT, Basdra EK, Papavassiliou AG. (2013) DNA methylation biomarkers as diagnostic and prognostic tools in colorectal cancer. *J Mol Med.* 91:1249–1256

70. Clarke MA, Wentzensen N, Mirabello L, Ghosh A, Wacholder S, Harari A, Lorincz A, Schiffman M, Burk RD. (2012) Human Papillomavirus DNA Methylation as a Potential Biomarker for Cervical Cancer *Cancer Epidemiol Biomarkers Prev.* 21:2125–2137
71. Frommer M, McDonald LE, Millar DS, Collis CM, Watt F, Grigg GW, Molloy PL, Paul CL. (1992) A genomic sequencing protocol that yields a positive display of 5-methylcytosine residues in individual DNA strands. *Proc Natl Acad Sci U S A.* 89:1827–1831
72. Rein T, Zorbas H DePamphilis ML. (1997) Active mammalian replication origins are associated with a high-density cluster of mCpG dinucleotides. *Mol Cell Biol.* 17:416–426
73. Acevedo LG, Sanz A, Jelinek MA. (2011) Novel DNA binding domain-based assays for detection of methylated and nonmethylated DNA. *Epigenomics.* 3:93–101
74. Hernández HG, Tse MY, Pang SC, Arboleda H, Forero DA. (2013) Optimizing methodologies for PCR-based DNA methylation analysis. *Biotechniques.* 55:181–197
75. Hiraoka D, Yoshida W, Abe K, Wakeda H, Hata K, Ikebukuro K. (2012) Development of a Method To Measure DNA Methylation Levels by Using Methyl CpG-Binding Protein and Luciferase-Fused Zinc Finger Protein. *Anal Chem.* 84:8259–8264
76. Tanaka K, Tainaka K, Kamei T, Okamoto A. (2007) Direct labeling of 5-methylcytosine and its applications. *J Am Chem Soc.* 129:5612–5620

77. Stains CI, Furman JL, Segal DJ, Ghosh I. (2006) Site-specific detection of DNA methylation utilizing mCpG-SEER. *J Am Chem Soc.* 128:9761–9765
78. Qin Y, Hurley LH. (2008) Structures, folding patterns, and functions of intramolecular DNA G-quadruplexes found in eukaryotic promoter regions. *Biochimie.* 90:1149–1171
79. Lin J, Hou JQ, Xiang HD, Yan YY, Gu YC, Tan JH, Li D, Gu LQ, Ou TM, Huang ZS. (2013) Stabilization of G-quadruplex DNA by C-5-methyl-cytosine in bcl-2 promoter: implications for epigenetic regulation. *Biochem Biophys Res Commun.* 433:368–373
80. Stevens AJ, Stuffrein-Roberts S, Cree SL, Gibb A, Miller AL, Doudney K, Aitchison A1, Eccles MR, Joyce PR, Filichev VV, Kennedy MA. (2014) G-Quadruplex Structures and CpG Methylation Cause Drop-Out of the Maternal Allele in Polymerase Chain Reaction Amplification of the Imprinted MEST Gene Promoter. *PLoS One.* 9:e113955
81. Gehring K, Leroy JL, Guéron M. (1993) A tetrameric DNA structure with protonated cytosine-cytosine base pairs. *Nature.* 363:561–565
82. Day HA, Pavlou P, Waller ZA. (2014) i-Motif DNA: structure, stability and targeting with ligands. *Bioorg Med Chem.* 22:4407–4418
83. Uribe DJ, Guo K, Shin YJ, Sun D. (2011) Heterogeneous Nuclear Ribonucleoprotein K and Nucleolin as Transcriptional Activators of the Vascular Endothelial Growth Factor Promoter through Interaction with Secondary DNA Structures. *Biochemistry.* 50:3796–3806

84. Yang B, Wu RR, Rodgers MT. (2013) Base-Pairing Energies of Proton-Bound Homodimers Determined by Guided Ion Beam Tandem Mass Spectrometry: Application to Cytosine and 5-Substituted Cytosines. *Anal Chem.* 85:11000–11006
85. Yang B, Moehlig AR, Frieler CE, Rodgers MT. (2015) Base-Pairing Energies of Protonated Nucleobase Pairs and Proton Affinities of 1-Methylated Cytosines: Model Systems for the Effects of the Sugar Moiety on the Stability of DNA i-Motif Conformations. *J Phys Chem B.* 119:1857–1868
86. Li G, Ruan X, Auerbach RK, Sandhu KS, Zheng M, Wang P, Poh HM, Goh Y, Lim J, Zhang J, Sim HS, Peh SQ, Mulawadi FH, Ong CT, Orlov YL, Hong S, Zhang Z, Landt S, Raha D, Euskirchen G, Wei CL, Ge W, Wang H, Davis C, Fisher-Aylor KI, Mortazavi A, Gerstein M, Gingeras T, Wold B, Sun Y, Fullwood MJ, Cheung E, Liu E, Sung WK, Snyder M, Ruan Y. (2012) Extensive promoter-centered chromatin interactions provide a topological basis for transcription regulation. *Cell.* 148:84–98
87. Yoshida W, Tomokawa J, Inaki M., Kimura H, Onodera M, Hata K, Nakabayashi K. (2015) An insulator element located at the cyclin B1 interacting protein 1 gene locus is highly conserved among mammalian species. *PLoS One.* 10:e0131204
88. Iida K, Nagasawa K. (2013) Macrocyclic polyoxazoles as G-quadruplex ligands. *Chem Rec.* 13:539–548
89. Tera M, Iida K, Ishizuka H, Takagi M, Suganuma M, Doi T, Shin-ya K, Nagasawa K. (2009) Synthesis of a potent G-quadruplex-binding macrocyclic heptaoxazole. *Chembiochem.* 10:431–435

90. Chung WJ, Heddi B, Tera M, Iida K, Nagasawa K, Phan AT. (2013) Solution structure of an intramolecular (3 + 1) human telomeric G-quadruplex bound to a telomestatin derivative. *J Am Chem Soc.* 135:13495–13501
91. Harada T, Iwai A, Miyazaki T. (2010) Identification of DELE, a novel DAP3-binding protein which is crucial for death receptor-mediated apoptosis induction. *Apoptosis.* 15:1247–1255
92. Feng CJ, Lu XW, Luo DY, Li HJ, Guo JB. (2013) Knockdown of Cdc6 inhibits proliferation of tongue squamous cell carcinoma Tca8113 cells. *Technol Cancer Res Treat.* 12:173–181
93. Zhang G, Liu K, Ling X, Wang Z, Zou P, Wang X, Gao J, Yin L, Zhang X, Liu J, Ao L, Cao J. (2016) DBP-induced endoplasmic reticulum stress in male germ cells causes autophagy, which has a cytoprotective role against apoptosis in vitro and in vivo. *Toxicol Lett.* 245:86–98
94. Rasheed MZ, Tabassum H, Parvez S. (2017) Mitochondrial permeability transition pore: a promising target for the treatment of Parkinson's disease. *Protoplasma.* 254:33–42
95. Zakraoui O, Marcinkiewicz C, Aloui Z, Othman H. (2017) Lebein, a snake venom disintegrin, suppresses human colon cancer cells proliferation and tumor-induced angiogenesis through cell cycle arrest, apoptosis induction and inhibition of *VEGF* expression. *Mol Carcinog.* 56:18–35

96. Hardin CC, Watson T, Corregan M, Bailey C. (1992) Cation-dependent transition between the quadruplex and Watson-Crick hairpin forms of d(CGCG3GCG). *Biochemistry*. 31:833–841
97. Fry M, Loeb LA. (1994) The fragile X syndrome d(CGG)<sub>n</sub> nucleotide repeats form a stable tetrahelical structure. *Proc Natl Acad Sci U S A*. 91:4950–4954
98. Zamiri B, Mirceta M, Bomszyk K, Macgregor RB Jr., Pearson CE. (2015) Quadruplex formation by both G-rich and C-rich DNA strands of the C9orf72 (GGGGCC)<sub>8</sub>•(GGCCCC)<sub>8</sub> repeat: effect of CpG methylation. *Nucleic Acids Res*. 43:10055–10064
99. Guo K, Gokhale V, Hurley LH, Sun D. (2008) Intramolecularly folded G-quadruplex and i-motif structures in the proximal promoter of the vascular endothelial growth factor gene. *Nucleic Acids Res*. 36:4598–4608
100. Guo K, Pourpak A, Beetz-Rogers K, Gokhale V, Sun D, Hurley LH. (2007) Formation of pseudosymmetrical G-quadruplex and i-motif structures in the proximal promoter region of the RET oncogene. *J Am Chem Soc*. 129:10220–10228
101. Seenisamy J, Rezler EM, Powell TJ, Tye D, Gokhale V, Joshi CS, Siddiqui-Jain A, Hurley LH. (2004) The dynamic character of the G-quadruplex element in the c-MYC promoter and modification by TMPyP4. *J Am Chem Soc*. 126:8702–8709
102. Nair SS, Coolen MW, Stirzaker C, Song JZ, Statham AL, Strbenac D, Robinson MD, Clark SJ. (2011) Comparison of methyl-DNA immunoprecipitation (MeDIP) and methyl-CpG binding domain (MBD) protein capture for genome-wide DNA methylation analysis reveal CpG sequence coverage bias. *Epigenetics*. 6:34–44

103. Chambers VS, Marsico G, Boutell JM, Di Antonio M, Smith GP, Balasubramanian S. (2015) High-throughput sequencing of DNA G-quadruplex structures in the human genome. *Nat Biotechnol.* 33:877–881
104. Zhang Y, Ozdemir P. (2009) Microfluidic DNA amplification--a review. *Anal Chim Acta.* 638:115–125
105. Aschenbrenner J, Drum M, Topal H, Wieland M, Marx A. (2014) Direct sensing of 5-methylcytosine by polymerase chain reaction. *Angew Chem Int Ed Engl.* 53:8154–8158

## Published papers

- Bay DH, Busch A, Lisdat F, Iida K, Ikebukuro K, Nagasawa K, Karube I, and Yoshida W. (2017) Identification of G-quadruplex structures that possess transcriptional regulating functions in the *Dele* and *Cdc6* CpG islands. *BMC Mol Biol.* 18:17
- Yoshida W, Yoshioka H, Bay DH, Iida K, Ikebukuro K, Nagasawa K, and Karube I. (2016) Detection of DNA Methylation of G-Quadruplex and i-Motif-Forming Sequences by Measuring the Initial Elongation Efficiency of Polymerase Chain Reaction. *Anal Chem.* 88:7101–7107