

AC-motif: a DNA motif containing adenine and cytosine repeat plays a role in gene regulation

Jeong Hwan Hur¹, Chan Young Kang², Sungjin Lee³, Nazia Parveen¹, Jihyeon Yu², Amen Shamim^{1,4}, Wanki Yoo¹, Ambarnil Ghosh¹, Sangsu Bae^{2,*}, Chin-Ju Park^{3,*} and Kyeong Kyu Kim^{1,5,*}

¹Department of Precision Medicine, Graduate School of Basic Medical Science (GSBMS), Institute for Anti-microbial Resistance Research and Therapeutics, Sungkyunkwan University School of Medicine, Suwon 16419, Republic of Korea, ²Department of Chemistry, Hanyang University, Seoul 04763, Republic of Korea, ³Department of Chemistry, Gwangju Institute of Science and Technology, Gwangju 61005, Republic of Korea, ⁴Centre of Agricultural Biochemistry and Biotechnology, University of Agriculture, Faisalabad 38040, Pakistan and ⁵Samsung Biomedical Research Institute, Samsung Advanced Institute for Health Sciences and Technology, Samsung Medical Center, Sungkyunkwan University School of Medicine, Seoul 06351, Republic of Korea

Received November 19, 2020; Revised August 04, 2021; Editorial Decision August 06, 2021; Accepted August 27, 2021

ABSTRACT

I-motif or C4 is a four-stranded DNA structure with a protonated cytosine:cytosine base pair (C⁺:C) found in cytosine-rich sequences. We have found that oligodeoxynucleotides containing adenine and cytosine repeats form a stable secondary structure at a physiological pH with magnesium ion, which is similar to i-motif structure, and have named this structure 'adenine:cytosine-motif (AC-motif)'. AC-motif contains C⁺:C base pairs intercalated with putative A⁺:C base pairs between protonated adenine and cytosine. By investigation of the AC-motif present in the CDKL3 promoter (AC-motif_{CDKL3}), one of AC-motifs found in the genome, we confirmed that AC-motif_{CDKL3} has a key role in regulating CDKL3 gene expression in response to magnesium. This is further supported by confirming that genome-edited mutant cell lines, lacking the AC-motif formation, lost this regulation effect. Our results verify that adenine-cytosine repeats commonly present in the genome can form a stable non-canonical secondary structure with a non-Watson–Crick base pair and have regulatory roles in cells, which expand non-canonical DNA repertoires.

INTRODUCTION

DNA can be folded into structures other than double stranded B-DNA, including Z-DNA, cruciform, triplex, hairpin, A-motif, guanine-quadruplex (G4) and i-motif (C4)

(1–6). It is known that these non-canonical structures participate in various biological functions, especially in maintaining genome integrity and regulating transcriptional activities (7–9). For instance, the structural and functional aspects of **G-quadruplex** have been intensively investigated to explain the roles of non-canonical DNA structures in cells (10). **G-quadruplex** was first identified in the human telomere region, where its formation was thought to affect the binding of certain proteins to maintain telomere integrity during meiosis (11). **G-quadruplex** is known to occur at various locations in the genome, especially at promoter regions of many oncogenes such as c-MYC (12), KRAS (13), BCL2 (14) and VEGF (15). Its formation and destabilization at the promoter region are known to affect gene transcription, eventually altering the proteome dynamics in different types of cancer cells.

i-motif can also be formed on the complementary strand where **G-quadruplex** can be formed in a mutually exclusive manner (16). Various evidences support that **i-motif** can exist in the nuclei of human cells, and it can also regulate the transcriptional level of certain genes upon formation or destabilization of its structure, especially in the promoter regions (17–19). In the **i-motif** structure, two sets of vertically orientated base pairs between hemi-protonated cytosine (C⁺) and cytosine (C) are intercalated to form a stable structure (20). Accordingly, **i-motif** is formed at acidic conditions since hemi-protonation at the N3 of cytosine is achieved at low pH and thus disintegrates due to breakage of the base pairs at physiological conditions. I-motif formations in molecular crowding condition as well as under negative super helical condition further supports the possibility of i-motif formation at physiological condition (21,22).

*To whom correspondence should be addressed. Tel: +82 31 299 6136; Fax: +82 31 299 6159; Email: kyeongkyu@skku.edu
Correspondence may also be addressed to Sangsu Bae. Email: sangsubae@hanyang.ac.kr
Correspondence may also be addressed to Chin-Ju Park. Email: cjpark@gist.ac.kr

Recently, the i-motif formation in the nuclei of human cells provided the evidence of which supports the existence of the **i-motif** structure in the genome and its roles in the near physiological surroundings (19).

As in the cases of **G-quadruplex** and **i-motif**, which are formed in repeated sequences, many repeated sequences whose structures and functions are not identified can possibly form non-canonical secondary structures in the genome. Considering the imperative roles of those non-canonical structures in regulating many genetic events, it is necessary to identify novel DNA motifs with repeated sequences and investigate their structures and functions for comprehensive understanding of the genome.

From this point of view, we investigated the role of adenine repeats, since their involvement in secondary structure formation is relatively unknown other than A-motif of which poly adenine repeats form parallel double strand helical structure via protonated adenine bases (6). First, we explored the effect of adenine repeats in combination with cytosine repeats by replacing cytosine repeats in the **i-motif** structure with adenine repeats in various combinations (Table 1) and testing their stabilities and structures using circular dichroism (CD), fluorescence and nuclear magnetic resonance (NMR) spectroscopic methods. By this approach, we identified a DNA motif containing adenine and cytosine repeats which forms a secondary structure, which resembles i-motif. We further confirmed that this new DNA motif present in the gene promoter region has a functional role in gene regulation using genome edited mutant cell lines and reporter assays. Therefore, we verified that adenine-cytosine repeats can form a secondary structure similar to i-motif with functional implications in gene regulation. We anticipate this current study not only contributes to nucleic acid chemistry by identifying a new type of non-canonical DNA and its role but also opens a new chapter in genome studies by presenting the possibility of discovering novel functional DNA motifs.

MATERIALS AND METHODS

Sample preparation

Stock sample. For the preparation of the DNA stock sample, oligodeoxynucleotides (ODNs) (Cosmogentech, Seoul, Republic of Korea) were purchased and dissolved in a storage buffer containing 90 mM NaCl, 5 mM sodium phosphate and 0.5 mM EDTA, pH 6.0. They were heated to 97°C, annealed slowly at room temperature over at least 4 h and stored overnight at 4°C. The annealed ODNs were concentrated to 1.5 mM using 3000 MW cut-off centricon (Merck Millipore, Burlington, MA, USA) at 4°C, and the concentrated sample (stock DNA sample) was stored in the storage buffer at -20°C before using. The concentration of DNA sample was measured using a NanoDrop (Thermo Fisher Scientific, Waltham, MA, USA).

CD sample. For the circular dichroism (CD) experiment, the stock DNA sample was diluted 100 times to prepare a CD sample (final 15 µM) in a CD buffer containing 5 mM sodium phosphate, pH 5.0–8.0, 5 mM cation such as MgCl₂, followed by heating to 97°C and annealing slowly at room temperature over at least 4 h. The CD sample was

Table 1. Oligonucleotides (ODNs) used in this study. i-motif_{hTelo} contains the i-motif-forming ODN present in the human telomere. ACR1–5 are the model ODNs containing trideoxyadenylate and trideoxycytidylate in various combinations connected by 3-nt linkers. ACR6–12 are the model ODNs containing trideoxyadenylate and trideoxycytidylate in various combinations connected by 6-nt linkers. exI-motif_{hTelo} contains four trideoxycytidylates connected by TCCTTT linkers. AC-motif_{CDKL3} contains the AC-motif-forming sequence present in the promoter region of CDKL3 gene. The mutant AC-motif_{CDKL3} has deoxyadenylates instead of deoxycytidylates in AC-motif_{CDKL3}. ACR13 and ACR14 have the same sequence as ACR6 except the linkers are TAAT and TAATT instead of TAATTT, respectively. The substituted sequences are underlined. The oligonucleotide sequences are written from 5' to 3', and deoxyadenylates and deoxycytidylates are highlighted in red.

Name	Sequence
i-motif _{hTelo}	5'- <u>CCCTAA</u> <u>CCCTA</u> <u>ACCCTA</u> <u>ACCC</u> -3'
exI-motif _{hTelo}	<u>CCCTAATTTCCCTAATTTCCCTAATTTCCC</u>
ACR1	<u>AAATAACCCCTAACCCCTAACCC</u>
ACR2	<u>AAATAACCCCTAAAAATAACCC</u>
ACR3	<u>AAATAACCCCTAACCCCTAAAA</u>
ACR4	<u>CCCTAAAAATAAAAAATAACCC</u>
ACR5	<u>CCCTAACCCCTAACCCCTAAAA</u>
ACR6	<u>AAATAATTTCCCTAATTTCCCTAATTTCCC</u>
(AC-motif)	
ACR7	<u>AAATAATTTCCCTAATTTAAATAATTTCCC</u>
ACR8	<u>AAATAATTTCCCTAATTTCCCTAATTTAAA</u>
ACR9	<u>CCCTAATTTAAATAATTTAAATAATTTCCC</u>
ACR10	<u>CCCTAATTTCCCTAATTTCCCTAATTTAAA</u>
ACR11	<u>CCCTAATTTAAATAATTTCCCTAATTTCCC</u>
ACR12	<u>CCCTAATTTCCCTAATTTAAATAATTTCCC</u>
ACR13	<u>AAATAATCCCTAATCCCTAATCCC</u>
ACR14	<u>AAATAATCCCTAATCCCTAATCCC</u>
Wild-type	<u>AAAAAAAGTCCCGCGCAGCCCCCCCACCC</u>
AC-motif _{CDKL3}	

used for the CD experiment immediately after annealing unless specified otherwise.

In the case of exploring concentration effect on CD experiments, the stock DNA sample was diluted 100 times (final 15 µM) in a CD buffer 2 (90 mM NaCl, 5 mM sodium phosphate at pH 5.0) for the purpose of buffer exchange. Then, it was concentrated to the concentration higher than 100 µM. The sample was serially diluted with the CD buffer2 to make 6, 15, 30, 60, 80 and 100 µM samples that were heated to 97°C, annealed slowly at room temperature over at least 4 h and stored overnight at 4°C followed by CD experiment.

For exploring the pH effect on CD experiments, the stock DNA sample was diluted 100 times (final 15 µM) in a CD buffer 2 with various pH (pH 5.0–8.0). The final CD samples were prepared in the same way as in the sample in CD buffer.

NMR sample. For the NMR experiment, an NMR sample was prepared in the similar way as the CD sample from the stock sample in an NMR buffer (5 mM sodium phosphate, pH 7.4, 5 mM MgCl₂). However, since the higher concentration is necessary for the NMR measurement, annealed NMR sample at the concentration of 15 µM in the NMR buffer was further concentrated. Therefore, the NMR sample having a DNA concentration of 0.5 mM was prepared in 500 µl of the NMR buffer and freeze-dried.

For 1D ^1H -NMR and 2D NOESY experiments, the freeze-dried (lyophilized) sample was dissolved in 500 μl 10% D_2O /90% H_2O and 100% D_2O , respectively, right before NMR experiments. The sample quality before and after lyophilization were confirmed by CD spectrometry (Supplementary Figure S9b).

Circular dichroism measurement

Circular dichroism (CD) experiment was performed using Jasco-810 (Jasco, Easton, MD, USA). The measurement was performed at wavelengths between 230 and 320 nm, with a 1 nm step width, and a 1 s response time to monitor the conformation of DNA. All presented CD spectra are the averaged spectra of three scans of the same sample with the baseline correction by considering the signal contribution from the buffer. The experiment was performed at 25°C. The ODN concentration used in all CD experiments was 15 μM in the final sample volume of 200 μl . CD spectra measured in mdeg per whole oligonucleotide were converted into molar extinction ($\Delta\epsilon$ in $\text{M}^{-1}\text{cm}^{-1}$) using the following equation (23).

1. $\text{mdeg} = A (\text{absorbance}) \times 32\,980$ (conversion factor)
2. $\epsilon = A \times M (\text{g/mol}) / (C \times L)$; M (molecular weight; g/mol), C (concentration; g/l), L (path length; cm)
3. $\theta [\text{mdeg}] = C \times L \times 32\,980 \times \epsilon / M$

Thermal melting analysis

Jasco-810 (Jasco, Easton, MD, USA) was used to perform thermal melting analysis to measure the stability of AC-motif. At a fixed wavelength of 283 nm, the temperature was increased from 10°C to 90°C with 2°C increment per minute. The thermal melting graph was analyzed using GraphPad software (GraphPad Software, La Jolla, CA, USA) upon applying the first derivative function. The processed graph was fit to find the zenith of the peak. The melting temperature (T_m) of ACR6 at each condition was calculated from the first derivative of CD value at 283 nm versus temperature.

Nuclear magnetic resonance (NMR)

NMR experiments were performed using either a Bruker 800 MHz or a Bruker 900 MHz spectrometer (Korea Basic Science Institute, Ochang Center, Republic of Korea). Both spectrometers are equipped with a cryogenic probe. All NMR samples were prepared at a concentration of 0.5 mM. 1D experiments were performed with the 11-echo pulse sequence (24) at three different temperatures (5.0, 15.0 and 25.0°C) (Korea Basic Science Institute, Ochang Center, Republic of Korea). 2D-NOESY spectra were acquired in H_2O with 300 ms mixing time at 5.0 and 25.0°C. We also performed 2D-NOESY experiments in D_2O with 300 ms mixing time at 25.0°C.

Molecular dynamics simulation

Molecular dynamics simulations were performed using the AMBER package (version 16) (25). The energy minimization and equilibration of the system and standard production simulations were analyzed with AMBER algorithm

using the AMBER force field. The tleap module of AMBER was used to prepare the simulations system. The Pmemd.cuda application was used to obtain the MD simulation trajectories while analyses were performed by CPPTRAJ module (26) and xmgrace (27). A pre-equilibrated elementary TIP3P tetrahedral box (28) was generated ~ 8 Å in each direction from the solute. The total volume of the box was 13 0607.600 Å³, and the overall compactness of the system was 0.787 g/cc. Twenty-three Na^+ counter ions were added to neutralize the system. All simulations were run with the SHAKE algorithm (29) to constrain hydrogen involving covalent bonds, with 2 fs periodic boundary conditions under 26.9°C (30). During 1000 ns simulation, analysis files were saved in every 0.5 ps.

CDKL3 promoter luciferase assay

A region 700 bases upstream of the transcription start site (TSS) and 300 bases downstream of the TSS promoter region was amplified from genomic DNA isolated from human red blood cells. The CDKL3 forward primer 5'-GGTACCAGCCGCCCTCTGAGG-3' with a KpnI site and the CDKL3 reverse primer 5'-AAGCTTGTGTGGGCTCAGTATTCAAG-3' with a HindIII site were used to obtain a 1 kb CDKL3 promoter region from genomic DNA. The PCR product was cloned into a pGL3-Basic luciferase vector. In order to induce mutation, the forward primer 5'-GATTCGGCCGCCGCGCCGGGGGTTTTTTT TCTGCG-CGGGACTTTTTTTTTT-3' and the reverse primer 5'-TGAAAAAAAAAAGTCCCGCCGC-GCAGAAAAAAAAAC-CCCCCGGCGCGGCGGCCG-3' were used. Each of 900 ng/ μl cloned firefly luciferase reporter vector and 100 ng/ μl of control Renilla luciferase vector were transfected to 5.0×10^5 cells/ml HEK293 cells at the same time using Turbofect (Thermo Fisher Scientific, Waltham, MA, USA). Magnesium-containing media were substituted after 24 h and cultured for another 24 h at 37°C. Cells were collected and washed with PBS followed by the usage of Dual-Luciferase® Reporter Assay kit (Promega Corporation, Madison, WI, USA) to measure luciferase activity. The fluorescence was measured using Tecan Infinite 200 PRO (Tecan, Männedorf, Switzerland).

Production of genome engineered cell lines

Transfection condition. Purified SpCas9 protein (15 μg) and single guide RNAs (sgRNAs) (10 μg) were incubated at RT to make ribonucleoprotein (RNP) complex, then transfected with 120-nt single-stranded oligodeoxynucleotides (ssODNs, 7.4 μg) as a donor template into 2×10^5 of HeLa cells using 4D-nucleofector (Lonza Group Ltd, Basel, Switzerland) 20 μl Nucleovette strip (Lonza Group Ltd, Basel, Switzerland) with CN-114 program. Genomic DNAs were isolated with the Nucleospin Tissue Kit (MACHERY-NAGEL, Düren, Germany) 72 h post transfection.

Targeted deep sequencing and data analysis. The target site was amplified using KOD Multi&epi (TOYOBO CO., LTD., Osaka, Japan) with adaptor primers. Miniseq system (Illumina, Inc., San Diego, CA, USA) was used for

deep sequencing with D501-508 and D701-D712 library. After Miniseq sequencing, paired-end reads were joined by Fastq-join. Then, Fastq-join files were analyzed with Cas-analyzer (<http://www.rgenome.net/cas-analyzer/>) (31) to derive knock-in frequency in bulk cells and single cell lines.

Producing knock-in single cell line. For single cell colony expansion, transfected cells were counted with hemocytometer and diluted to a concentration of 1 cell/well. Then spread them into 96-well plates. After 10 days of incubation, only single cell colonies were screened by microscope. Single cell colonies with >70% confluence were shifted to 24-well plates. Genomic DNA of each single colony was isolated, and DNA sequences near the target region of CDKL3 were analyzed by targeted deep sequencing and also confirmed by Sanger sequencing (Supplementary Figure S14).

Genome wide mapping of AC-motif

Genome wide analyses of putative AC-motif-forming sequences in the human genome (hg37) were conducted by applying an in-house script written in R software (<https://www.r-project.org/>). The sequences containing the following patterns were assigned as putative AC-motif-forming sequences: A₃N₆C₃N₆C₃N₆C₃ or C₃N₆C₃N₆C₃N₆A₃, where A, C and N represent adenine, cytosine and any base including cytosine and adenine, respectively.

Quantitative real-time PCR

Total RNA was isolated using Trizol (Thermo Fisher Scientific, Waltham, MA, USA) followed by chloroform treatment. The aqueous phase was carefully isolated and precipitated with 100% isopropanol. The pellet was centrifuged and washed with 70% ethanol. About 2.5 µg of RNA was used for each RT reaction and reversely transcribed to cDNA using EcoDry™ Premix (Oligo dT) kit (Takara Bio USA, Inc., Mountain View, CA, USA). The reaction was performed under 42°C for 60 min followed by reaction quenching under 70°C for 10 min. SYBR Green qRT-PCR 2X premix (Bio-rad, Hercules, CA, USA) was used for the RT-PCR reaction using a CFX Connect machine (Bio-rad, Hercules, CA, USA). All samples were run in triplicate. Transcript levels were normalized to the Gapdh expression. Primers used are listed below.

Gapdh [NM_001357943.2]: TGCACCACCAACTGCTTAGC, GGCATGGACTGTGGTCATGAG
Cdk13 [NM_001113575.2]: GCACGAACACTAGCAGCTC, GATGGGTGGCATTGTGCACAG
C19orf73 [NM_018111.3]: TGTGTTTGGGAAGGTGGAGTG, TCCTCCAGAGTCTGAAACG

Western blot analysis

Wild-type and genome-engineered mutant HeLa cells were cultured in 6-well plates. A total of 5.0×10^5 cells were cultured in DMEM growth media (WELGENE Inc., Gyeongsangbuk-do, Republic of Korea) without Mg²⁺ for 24 h. The media were changed with Mg²⁺-containing media after 24 h of culture and collected using 2× sample buffer

after another 24 h of culture. The samples were boiled and loaded into the gel. The proteins were transferred to the nitrocellulose (NC) membrane and blocked with 5% skim milk. The membranes were washed with 0.05% TBS-T (Tris-buffered saline with 0.05% tween-20) several times. CDKL3 or GAPDH antibodies (Cusabio Biotech Co, Houston, TX, USA) were incubated overnight and detected with specific HRP-conjugated secondary antibody. Their signals were developed on X-ray film (Agfa-Gevaert N.V., Mortsel, Belgium).

Fluorescence resonance energy transfer (FRET) analysis

5' FAM and 3' BHQ-1 labeled oligonucleotides for ACR6 (AC-motif), 1A2C and 3CCC were purchased and dissolved in 90 mM NaCl, 5 mM NaPO₄ and 0.5 mM EDTA, pH 6.0. About 15 µM of labeled ODNs were used for the experiment. The light was excited at 495 nm and the fluoresced emission was monitored at the range from 505 to 650 nm. The λ_{max} was observed at 525 nm. The baseline of buffer was subtracted from each data accordingly.

Statistical analysis

Statistical significance was assessed by two-tailed *t*-test using GraphPad software (GraphPad software, La Jolla, CA, USA).

RESULTS

A novel DNA motif that forms the secondary structure

Although many adenine repeats are found in the genome, not many adenine-repeat DNA structures have been reported, though guanine and cytosine repeats are known to be involved in forming a stable DNA structure. Therefore, we hypothesized that an adenine repeat can also be part of a stable DNA structure. To test this hypothesis, we made model oligodeoxynucleotides (ODNs) that contain adenine and cytosine repeats by exchanging three adenine nucleotides with the cytosine repeats in various combinations (ACR1–ACR5, Table 1) using a cytosine-repeat sequence present in the human telomere, d(CCCCTAA)₃CCC (**i-motif_{hTelo}**, Table 1), as a template. In these model ODNs, different combinations of cytosine or adenine repeats are connected by a linker with the three nucleotides TAA. To release the possible structural restraints imposed by the short linker between cytosine or adenine repeats, we generated other model sets, ACR6–ACR14 and **exI-motif_{hTelo}** ACR6–ACR14, Table 1), by increasing the length of the linker nucleotides of ACR1–ACR5 and **i-motif_{hTelo}** to TAATTT. We then examined their potential to form a secondary structure at various conditions by circular dichroism (CD) spectroscopy since CD is commonly used for monitoring the secondary structures of nucleic acids (Supplementary Figures S1 and S2). The CD spectra of **i-motif_{hTelo}** and **exI-motif_{hTelo}** were used as controls for ACR1–ACR5 and ACR6–ACR10, respectively. **i-motif_{hTelo}** showed a maximum positive peak at 288 nm and minimum negative peak at 257 nm at acidic conditions, but the maximum positive peak was blue shifted to 275 nm with $3.3 \times 10^3 \text{ M}^{-1}\text{cm}^{-1}$ lower in molar extinction (ε) at a higher pH (Supplementary Figure S1A). **exI-motif_{hTelo}** showed a maximum positive peak at 283 nm and

minimum negative peak at 250 nm at acidic conditions, but the maximum peak shifted to 275 nm with $6.3 \times 10^3 \text{ M}^{-1}\text{cm}^{-1}$ lower in intensity at a higher pH (Figure 1A). These results suggest **i-motif**_{hTelo} and **exI-motif**_{hTelo} can form secondary structures only at acidic conditions and lose their secondary structure at a higher pH, which is well consistent to the previously known typical structural feature of **i-motif**. While **ACR1–ACR5** did not show any spectral signatures representing the CD spectra of the **i-motif** structure at any given conditions (Supplementary Figure S1A), **ACR6** showed the CD spectrum that is similar to the typical CD spectra of **i-motif** at pH 5.0, and thus we named the secondary structure ACR6 as an **i-motif-like** structure based on their CD spectra. However, the CD spectra of ACR6 varied when pH was elevated from pH 6.0 to 8.0 (Figure 1B). Although **ACR10** also displayed a spectral pattern similar to **ACR6** (Supplementary Figure S1B), we used **ACR6** for further structural study due to its higher CD value.

We further tested the effect of divalent cations on the secondary structure formation by examining the CD spectra in the presence of Ca^{2+} , Cd^{2+} , Co^{2+} , Mg^{2+} , Ni^{2+} , Zn^{2+} and Mn^{2+} at pH 5.0 (Figure 1C and Supplementary Figure S2), since various cations are known to stabilize nucleic acid structures (32,33). While most of metals reduced the CD ellipticities of ACR6 at 283 nm, Ca^{2+} and Mg^{2+} slightly increased CD value (Figure 1C). Therefore, we hypothesized that these metals might contribute to the formation or stability of the ACR6 structure, and thus we further tested the effect of Ca^{2+} or Mg^{2+} on ACR6 at pH 8.0 (Figure 1D). Interestingly, we found ACR6 showed **i-motif-like** CD spectra even at pH 8.0 in the presence of these metals. Previously, it was known that **i-motif** structure can be further stabilized in the presence of Ca^{2+} or Mg^{2+} along with 40% polyethylene glycol (PEG) 200 (34) at the neutral pH (7.0). However, current CD experiments revealed that ODNs containing adenine and cytosine repeats can form an **i-motif-like** secondary structure at the basic pH in the presence of Ca^{2+} or Mg^{2+} even without PEG. Although Ca^{2+} had a stabilization effect on ACR6 at the higher pH, we further examined the structural and functional implication of ACR6 in the presence of Mg^{2+} since CD value of ACR6 was higher in the condition of Mg^{2+} at pH 8.0 (Figure 1D). A titration experiment of Mg^{2+} concentration ranging from 15 μM to 15 mM (Supplementary Figure S3) revealed that **i-motif-like** secondary structure is well maintained in the presence of Mg^{2+} at concentrations $>1.5 \text{ mM}$. However, the structure seems to partially unfold at Mg^{2+} concentrations lower than 1.0 mM, since the maximum peak value was blue-shifted from 283 to 276 nm and CD value showed $3.3 \times 10^3 \text{ M}^{-1}\text{cm}^{-1}$ decrease. This result implies that a certain level of Mg^{2+} is necessary for the formation of **i-motif-like** secondary structure at the higher pH.

Since the current CD results revealed that the linker length may be a variable for DNA structure formation, we further tested the effect of the linker length on the structure formation by monitoring CD spectra of ODNs containing linkers of three to five (**ACR1**, **ACR13**, and **ACR14**) nucleotides between A- or C-repeats (Table 1, Supplementary Figure S2A and S4). From this study, it was confirmed that an **i-motif-like** secondary structure can be formed when the linker is longer than four nucleotides. Next, the thermal

stability of ACR6 was investigated by CD melting analysis in the presence of Mg^{2+} at both pH 5.0 and 8.0 within temperature ranges from 10°C to 90°C (Supplementary Figure S5). We found that T_m values of ACR6 at pH 5.0 and 8.0 in the presence of magnesium are the same (42.5°C) but lower than that of **i-motif** (53.0°C) (Supplementary Figure S5A). Furthermore, we measured the CD spectra and T_m of ACR6 at various concentrations ranging from 6 to 100 μM , in order to examine the concentration-dependent CD value and T_m change, which possibly represent the presence of intermolecular structures. Although CD value changed in concentration dependent manner, there was no T_m change under the tested temperature conditions (Supplementary Figure S5B and C). Therefore, it can be concluded that ACR6 mostly forms intramolecular structure.

We have tested the ODNs that have internal adenine repeats in other slots: ACR11 (with adenine repeat in the second slot) and ACR12 (with adenine repeat in the third slot). Interestingly, both ACR11 and ACR12 can form **i-motif-like** structures in the presence of Mg^{2+} ion at both pH 5.0 and 8.0 (Supplementary Figure S6A and B). This result shows that the ODNs having an adenine repeat at any of four slots can form **i-motif-like** structure in the presence of Mg^{2+} ion at low and high pH conditions.

To further elucidate the overall folding of ACR6 by confirming the orientation between 5' and 3' ends, we conducted a FRET experiment using ACR6 containing a FAM at its 5' end and its corresponding quencher BHQ-1 at its 3' end (Supplementary Figure S7A). When pH was increased from 4.0 to 9.0 in the absence of Mg^{2+} ion, the fluorescence intensity at 525 nm (FAM) of ACR6 also increased (Supplementary Figure S7B and C) due to the fluorescing of FAM signal, suggesting that 5' and 3' ends are closely located at lower pH but their relative distance increases at higher pH. However, in the presence of Mg^{2+} , we observed that the fluorescence signal of FAM conjugated to ACR6 did not fluoresced as pH increased from 4.0 to 9.0, suggesting that the **i-motif-like** structure are well maintained even at the higher pH in the presence of Mg^{2+} (Supplementary Figure S7C and D). These results can be interpreted by that **i-motif-like** structure forms in pH- and metal-dependent manners, which are consistent with the CD results.

Formation of an **i-motif-like** secondary structure via hemi-protonated adenine and cytosine base pairs

Based on the CD analysis, we confirmed ACR6 has the **i-motif-like** secondary structure. The canonical **i-motif** structure is stabilized by the hemi-protonated cytosine (C^+) and cytosine (C) base pairs in the opposite strands. Therefore, the question is how the adenine repeats contribute to the secondary structure formation. To investigate the secondary structure and base interactions in ACR6, we performed 1D ^1H -NMR experiments of ACR6 at pH 8.0 in the presence of Mg^{2+} at various temperatures (25, 15 and 5°C) (Supplementary Figure S8). From the 1D ^1H -NMR spectra of ACR6, we found imino peaks at 15.3–15.5, 13.5–13.8 and 10.3–11.0 (Supplementary Figure S8). We assigned the peak at 15.3–15.5 ppm to the resonance of imino protons of the hemi-protonated cytosines ([C]N3H) based on the previous NMR studies of **i-motif** (5). Peaks at 10.3–11.0 ppm are

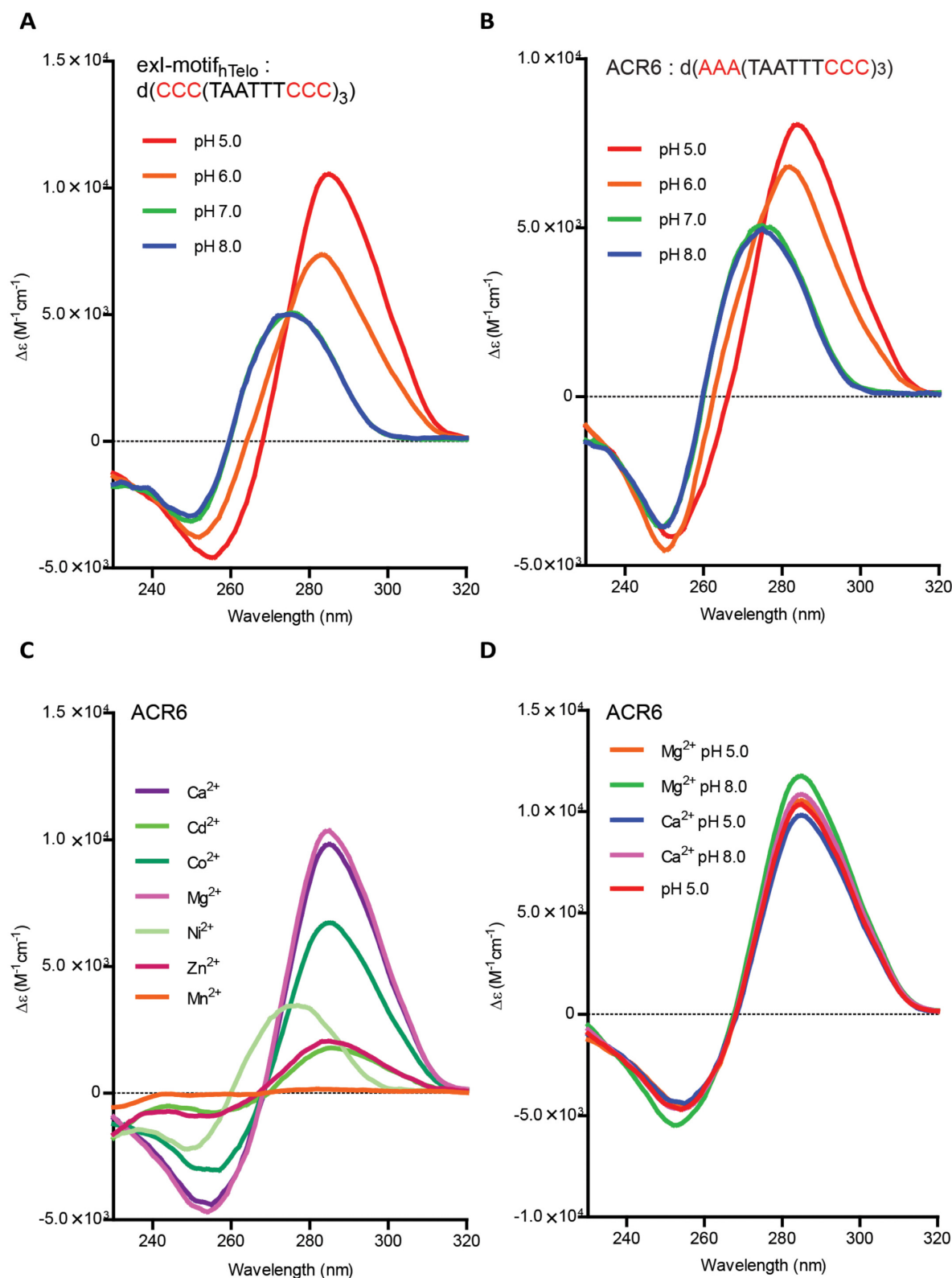


Figure 1. Circular dichroism (CD) spectra of ACR6 under various pH and metal conditions. (A) The CD spectra of 15 μ M exI-motif_{hTelo} in 90 mM NaCl, 5 mM NaH₂PO₄ at pH 5.0, 6.0, 7.0 and 8.0. (B) The CD spectra of 15 μ M ACR6 in 90 mM NaCl, 5 mM NaH₂PO₄ at pH 5.0, 6.0, 7.0 and 8.0. (C) The CD spectra of 15 μ M ACR6 in the presence of various divalent cations at an acidic condition (pH 5.0): 5 mM of Ca²⁺, Cd²⁺, Co²⁺, Mg²⁺, Ni²⁺, Zn²⁺ and Mn²⁺. (D) The CD spectra of 15 μ M ACR6 in the presence of 5 mM Mg²⁺ and Ca²⁺ at pH 5.0 and 8.0, respectively. The CD spectrum of 15 μ M exI-motif_{hTelo} in 90 mM NaCl, 5 mM NaH₂PO₄ at pH 5.0 was used as a control.

well matched to the peaks of the typical N3 imino proton of thymine in the absence of H-bonds ([T]N3H) (5,35). Peaks at 13.5–13.8 ppm became more evident and distinguishable at 5.0°C (Supplementary Figure S8), representing the presence of the temperature-dependent line broadening. These peaks can be interpreted as either [T]N3H imino proton in the Watson–Crick (WC) base pair (36) or imino proton similar to imino proton at the adenine N1 nitrogen reported in the c-GAMP riboswitches (37). Considering the geometric constraints in the loop of canonical i-motif, A:T base pair may not be present in the loop of ACR6 (38). Furthermore, if 13.5–13.8 ppm peaks correspond to [T]N3H from WC base pairs, they would be more stable at room temperature (25°C) than what we have observed in 1D ¹H-NMR spectra (Supplementary Figure S8). Therefore, we interpreted that those proton resonances at 13.5–13.8 ppm are originated from the imino proton from the hemi-protonated adenine ([A]NH⁺).

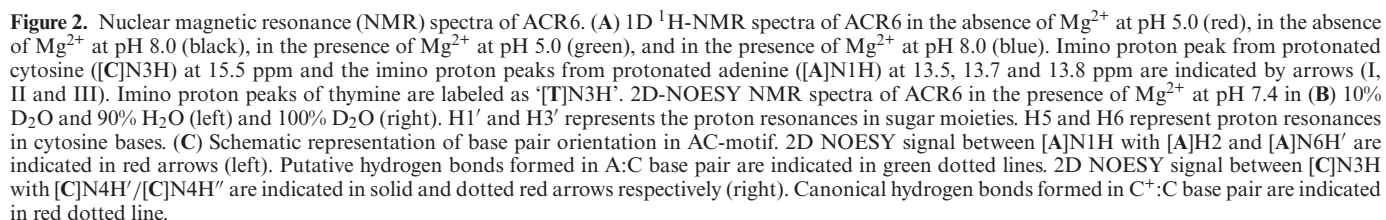
To further support this interpretation, we performed NMR analysis of ACR6 at varying pH and metal conditions (Figure 2A). We found that imino proton peaks at 15.3, 13.8, 13.6 and 13.5 ppm under the condition of pH 8.0 in the presence of Mg²⁺ were also observed at the lower pH regardless of the presence of Mg²⁺ but either disappeared or altered in the absence of Mg²⁺ at pH 8.0 (Figure 2A). This metal-dependent and pH-independent spectral observation is consistent with the previous CD and FRET results, suggesting that peaks at 15.3 ppm and 13.5–13.8 ppm are highly likely to be the imino proton peaks present in the i-motif-like structure of ACR6. Resonance at 15.3 ppm suggests the protonation at N3 position of cytosine ([C]N3H). However, it is not clear which nitrogen atoms in adenine are protonated from the current results; N1 and N7 in adenine are the possible protonation site. In order to elaborate on the identity of imino proton resonance and tertiary structure information of ACR6, a 2D-NOESY experiment was conducted using ACR6 at pH 7.4 with Mg²⁺ in 10% D₂O (Figure 2B, left) and 100% D₂O (Figure 2B, right). Unfortunately, the sequence specific resonance assignments were not feasible due to the overlapping of signals. However, we were able to identify some expected cross-peaks based on the previously known assigned peaks in other nucleic acid secondary structures (5,35–39). We found strong cross-peaks between resonance in the range of 7.0–7.2 ppm and resonances in the range of 13.5–13.8 ppm (Box a, Figure 2B). Since the peak at 7.0–7.2 ppm indicates the proton resonance of hydrogen at N6 of adenine ([A]N6H') (39), the cross-peaks shown in 'Box a' represent the proximity between the imino proton ([A]NH⁺) of adenine and hydrogen at N6 of adenine. We also observed cross-peaks between 7.6 ppm and 13.5–13.8 ppm (Box a, Figure 2B), suggesting that [A]NH⁺ and [A]H2 located closely since [A]H2 shows resonance at 7.6 ppm (40). Based on the observation that the protonation site is proximal to [A]N6H' and [A]H2 at pH 7.4, the N1 position of adenine ([A]N1) is highly likely to be protonated (Figure 2B).

Since the proton signals at 13.5–13.8 ppm become evident at low temperature due to slow exchange rate to the bulk solvent (Supplementary Figure S8), it is expected that the proton at N1 position of adenine ([A]N1H) is not exposed to the bulk solvent but rather it is involved in base

pairing in higher order DNA structure to further achieve the stability of the proton at N1 position. Therefore, we proposed that hemi-protonated adenine and cytosine form an A⁺:C base pair through [A]N1-H⁺-[C]N3 hydrogen bond as shown in Figure 2C (left). In this proposed model, hydrogen bonds [A]N6H'-[C]O2 and [A]N1H-[C]N3 are also expected (Figure 2C, left). Although we can confirm the intra-base cross peaks in adenine ([A]N6H'-[A]N1H and [A]H2-[A]N1H), we cannot assign the inter-base cross peaks such as [A]N1H-[C]N4H' possibly due to the weak intensity and overlap of peaks near 13.5/7.3 and 13.8/7.3 ppm in 'Box a' (Figure 2B). In addition to the A⁺:C base pair, C⁺:C base pair (Figure 2C, right) can be also present within i-motif like structure (ACR6) in the presence of Mg²⁺ at pH 7.4 considering the imino peaks in 1D NMR (Figure 2A) and 2D NMR peaks (Figure 2B): 'Box b' represents the cross-peaks of the protonated cytosine at N3 ([C]N3H:15.5 ppm) with a pair of hydrogen ([C]N4H':9.0 ppm and [C]N4H':7.3/8.0 ppm) at hydrogen bonded (int) and non-hydrogen bonded (ext) cytosine respectively (Box b, Figure 2B, left), which is also a typical characteristic of i-motif. While both inter- and intra-base pair cross peaks are observed in the C⁺:C base pair, only intra-base cross peaks are assigned in the case of the A⁺:C base pair (Box a, Figure 2B), suggesting that A⁺:C base pair is weaker than C⁺:C base pair. To further validate the current interpretation, we obtained the 2D-NOESY spectra from ACR6 at the acidic condition (pH 5.0) in the absence of Mg²⁺. Since the same i-motif-like structure can be formed in the absence of metal ions at the acidic condition (pH 5.0) according to the CD and fluorescence experiments, the similar pattern of 2D spectra as those obtained at the higher pH in the presence of Mg²⁺ is expected. Indeed, the similar cross-peaks were also observed at the acidic condition (Supplementary Figure S9).

Next, we analyzed sugar orientation of the base pairs in ACR6 by 2D NOESY signal obtained in 100% D₂O condition (Figure 2B, right). We found highly clustered cross-peaks at around 6.0 ppm (Box c, Figure 2B), which are distinctively found in the typical i-motif structure (5). These cross-peaks represent the close proximity of sugar hydrogens (H1') across the minor groove. In addition, the cross-peaks shown in 'Box d' represents the close proximity between H1' at around 6.0 ppm of sugar and H6 at around 7.1–7.7 ppm of cytidine and the strong cross-peaks shown in 'Box e' represent the close proximity between the intra-nucleotide H3' (4.5–5.0 ppm) and H6 (7.1–7.7 ppm) of cytidine. These supports 3'-endo sugar puckering of bases in ACR6, which are well reported in the typical i-motif structure (5,41). In addition, in the 2D-NOESY NMR spectra obtained in 10% D₂O and 90% H₂O (Figure 2B, left), the close proximity between H3' (4.5–5.0 ppm) of sugar and CN4H' (ext) (7.3/8.0 ppm) of cytidine is also expected by the cross peak in the 'Box f'. These results together reveal that the overall sugar conformation in ACR6 contains 3'-endo sugar pucker that is similar to that of the typical i-motif structure (5).

Our results demonstrate that ACR6 has an i-motif-like 3D structure at both acidic and physiological conditions in the presence of Mg²⁺ by sharing the following structural features with i-motif; C⁺:C base pairings, intercalation of base pairs, sugar orientation, and sugar pucker with a



3'-endo conformation, and global anti-parallel conformation of structure that resemble canonical i-motif structure (41,42). The difference between i-motif and ACR6 structure is the existence of putative A⁺:C non-canonical base pairs (Figure 2C, left, green dotted line) that are intercalated with C⁺:C base pairs (Figure 2C, right). Thus, we named ACR6 'adenine-cytosine motif (AC-motif).

Based on the proposed 3D-structure of ACR6 from the CD, fluorescence and NMR spectra, we made a model of ACR6 using the known four-stranded i-motif structure (PDB id: 1BQJ) (43) as a template by replacing the cytosine with adenine in the first strand of i-motif (Figure 3A). In this AC-motif model, hemi-protonated adenine and cytosine base pairs are intercalated with hemi-protonated cytosine and cytosine base pairs. In order to validate our model and test its stability, we performed a molecular dynamics (MD) simulation of the AC-motif structure for 1000 nanoseconds (ns) with the i-motif structure (PDB id: 1BQJ) (43) as a control (Figure 3B). For comparison, we also performed the MD simulation of AC-motif and i-motif models without protonation (Supplementary Figure S10A). By this simulation, we confirmed that the protonation at the [C]N3 for C⁺:C and [A]N1 for A⁺:C are necessary for stabilization of the structure since MD simulation model without protonation showed large root mean square deviation (RMSD) possibly due to the lack of C⁺:C and A⁺:C base pairs (Supplementary Figure S10A). Although it was noteworthy that AC-motif is less stable than i-motif as demonstrated by the relatively higher RMSD, both AC-motif and i-motif structures with protonation showed lower RMSD fluctuation (Figure 3B) than deprotonated models (Supplementary Figure S10A).

Identification of AC-motif in the human genome

To elucidate the biological abundance and functional implication of AC-motif, putative AC-motif-forming sequences were predicted in the human genome. We searched the AC-motif-forming sequence based on the scheme A₃N₆C₃N₆C₃N₆C₃, where N can be any nucleotide including adenine and cytosine. With this analysis, we identified 2151 putative AC-motif-forming sequences in the entire human genome. In addition, we also found that they are widely distributed in each chromosome with different numbers: from 16 AC-motifs in chromosome Y to 171 AC-motifs in chromosome 1 (Supplementary Figure S11). Furthermore, we searched for the putative AC-motifs with a linker length of 4 or 5 in the genome. Accordingly, we found 4397 AC-motifs with a linker length of 4, and 4572 AC-motifs with a linker length of 5 (Supplementary Figure S11). We then chose 77 putative AC-motif-forming sequences present in the gene promoter regions based on the information in the eukaryotic promoter database (EPD) (44) to study the functional implication of AC-motif in gene regulation (Supplementary Table S1). From the CD analysis of 77 ODNs covering the selected AC-motif-forming sequences present in the promoter, respectively, we learned that most ODNs with putative AC-motif-forming sequences showed similar CD spectra to ACR6 in the presence of Mg²⁺ at pH 8.0 (Supplementary Figure S12), suggesting that all predicted AC-motif-forming sequences in promoter

regions of the human genome indeed form a stable AC-motif structure.

The role of CDKL3 AC-motif in transcriptional activity

Among the 77 AC-motif candidates found in various human gene promoters, AC-motif_{CDKL3}, d(AAAAAAAGTCCCGCGCAGCCCCCCCCACCCC), 490 bp upstream of the transcription start site (TSS) of cyclin dependent kinase like 3 (CDKL3) gene (Figure 4A), was selected for structural and functional studies. Since CDKL3 is known to have cell proliferative effects in anchorage-independent HeLa cells (45), it was expected that its investigation would provide the biological implication of AC-motif in cell proliferation and possibly cancers as well. As a control, we tested a mutant AC-motif_{CDKL3} (Mut-AC-motif_{CDKL3}) that has adenine nucleotides in the position of the third cytosine run (III) of AC-motif_{CDKL3} (Figure 4A). While the CD spectra of AC-motif_{CDKL3} displayed the positive peak at 286 nm with CD value of $7.2 \times 10^6 \text{ M}^{-1}\text{cm}^{-1}$ in the presence of Mg²⁺ at both low and high pH (Figure 4B), Mut-AC-motif_{CDKL3} showed significantly lower CD value $3.3 \times 10^6 \text{ M}^{-1}\text{cm}^{-1}$ as well as a slight blue wavelength shift even in the presence of Mg²⁺ (Figure 4B). Therefore, it is certain that Mut-AC-motif_{CDKL3} cannot form a completely folded AC-motif structure even in the presence of Mg²⁺. Our CD results suggest that AC-motif_{CDKL3} maintains the stable i-motif-like structure (AC-motif) at physiological pH in the presence of Mg²⁺.

To test the effect of AC-motif_{CDKL3} on transcription activity, a 1 kb promoter region covering -700 to +300 nucleotides with TSS as a reference point was cloned into the luciferase reporter vector pGL3-basic (pAC-motif_{CDKL3}; Figure 4C), and the transcription activity was examined by measuring the Firefly luciferase activity relative to Renilla luciferase activity in HEK293 cells in the presence of varied amounts of Mg²⁺ (Figure 4D). As a control, the 1 kb promoter region containing Mut-AC-motif_{CDKL3} was inserted in the same reporter vector and activity was monitored (pMut-AC-motif_{CDKL3}). While the luciferase activity of pAC-motif_{CDKL3} showed a gradual increase upon the enhanced concentration of Mg²⁺, the luciferase activity of pMut-AC-motif_{CDKL3} remained unchanged as the concentration of Mg²⁺ in the media increased from 0 to 10 mM (Figure 4D). This suggests that Mg²⁺-dependent AC-motif formation is relevant to the regulation of luciferase gene expression. In order to confirm the influx of magnesium ion from the extracellular environment to the cell, we stained the cell with Magnesium GreenTM (Thermo Fisher Scientific), which emits a green fluorescence signal upon metabolizing the magnesium ion inside the cells. By observing the increased fluorescence intensity of the dye, we confirmed that the concentration of the intracellular magnesium ion increased in HeLa cells upon addition of magnesium in the media (Supplementary Figure S13).

Effects of AC-motif formation on the expression of endogenous CDKL3

To provide direct evidence supporting the regulatory effect of AC-motif_{CDKL3} on gene expression at a cellular level, stable mutant HeLa cell lines containing

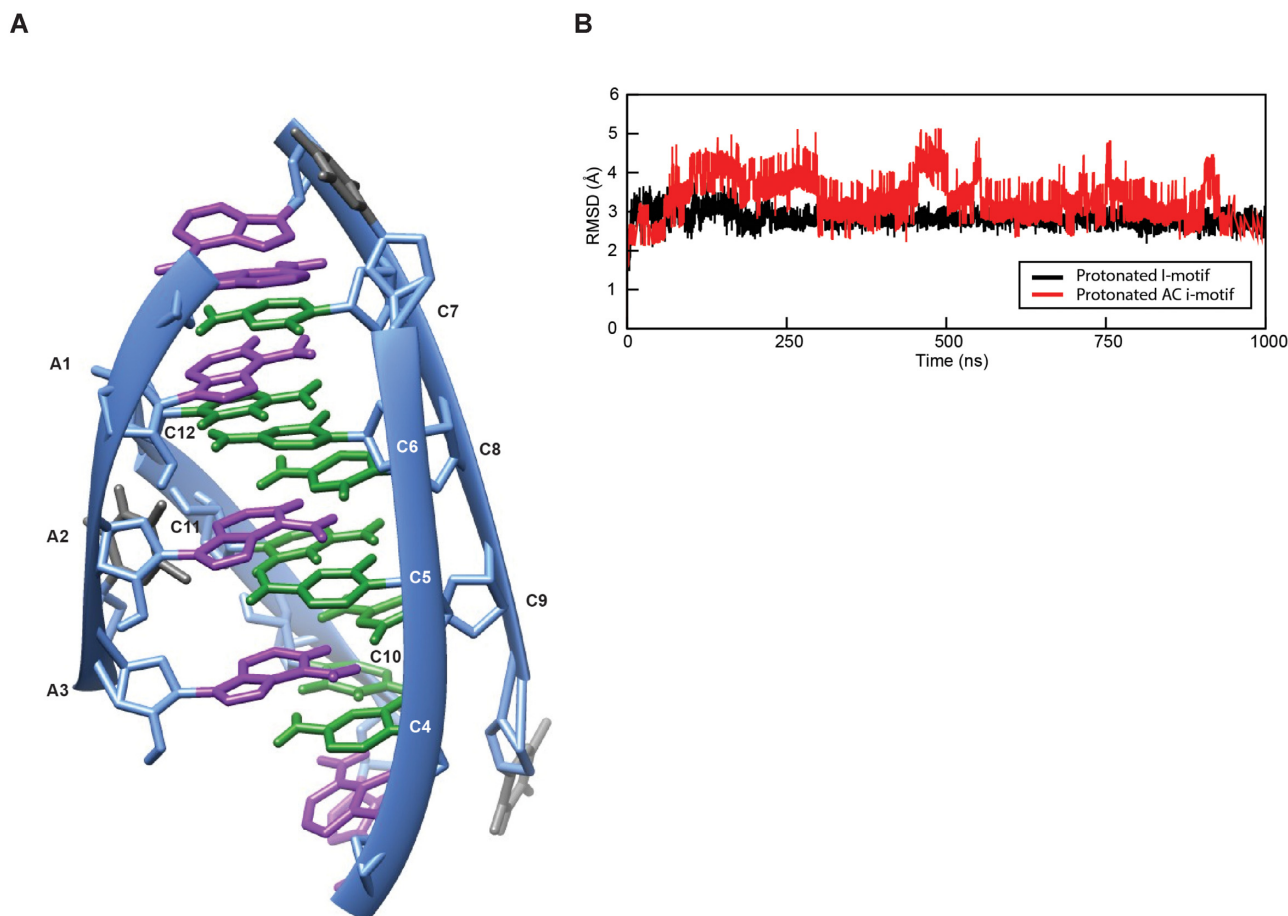


Figure 3. Molecular dynamic simulation of AC-motif and i-motif models. (A) Three-dimensional structure of the AC-motif model. The backbone (blue) and bases (green for cytosine and magenta for adenine) are represented in ribbon and stick models, respectively. (B) Graphical representation of the root mean square deviation (RMSD) of atoms in the protonated AC-motif (red) and i-motif (black) models during 1000 ns molecular dynamics simulation.

Mut-AC-motif_{CDKL3} were generated by CRISPR-Cas9 (Clustered Regularly Interspaced Short Palindromic Repeats and CRISPR-associated protein 9) technique (Figure 5a and Supplementary Figure S14) and the endogenous CDKL3 level was examined in response to the magnesium ions. By evaluating cell viability of each cell line in the presence of Mg^{2+} , we found that >50% cells can survive in the growth media containing 10 mM Mg^{2+} (Figure 5B), implying that the cell assay in the presence of <10 mM of Mg^{2+} is valid. In the wild-type cell lines, mRNA levels of *Cdkl3* increased up to 14.2-fold when Mg^{2+} concentration in the growth media was increased to 10 mM (Figure 5C). However, the mRNA expression level decreased 4.2-fold by Mg^{2+} treatment in the mutant cell line (Figure 5C). Since the secondary structure of G-rich sequence on the strand complementary to **AC-motif_{CDKL3}** was not affected by Mg^{2+} at pH 8.0, we were able to omit the possibility of an effect from the G-quadruplex on the transcription of CDKL3 (Supplementary Figure S15). When endogenous CDKL3 protein levels were monitored by western blotting, we observed a similar metal-dependent expression pattern (Figure 5D), although the fold difference was different from the mRNA results. Together, these results imply that **AC-motif** formation in the promoter region affects the transcription of CDKL3,

subsequently affecting protein levels. To further examine the role of **AC-motif** in controlling the gene expression, we additionally examined the transcription of the gene that has AC-motif in the promoter region. From the genome analysis (Supplementary Table S1), we selected C19orf73 as a candidate. We have tested its mRNA increment as we increased the concentration of Mg^{2+} (0–10 mM) in the growth media. In our result, we consistently observed that mRNA expression increased in a concentration-dependent way when Mg^{2+} was treated (Supplementary Figure S16).

DISCUSSION

In this study, we verified that ODNs containing adenine and cytosine repeats form a stable **i-motif-like** structure at the physiological pH in the presence of Mg^{2+} , and that they have a biological role in controlling transcription in a metal-dependent manner (Figures 1B–D and 6). Although several non-canonical DNAs with important cellular activities have been identified and studied, it has been hypothesized that more DNA motifs with repeated sequences could form stable secondary structures with various functionalities. We provide evidences supporting this hypothesis by proving the presence and functionality of **AC-motif**. In structural

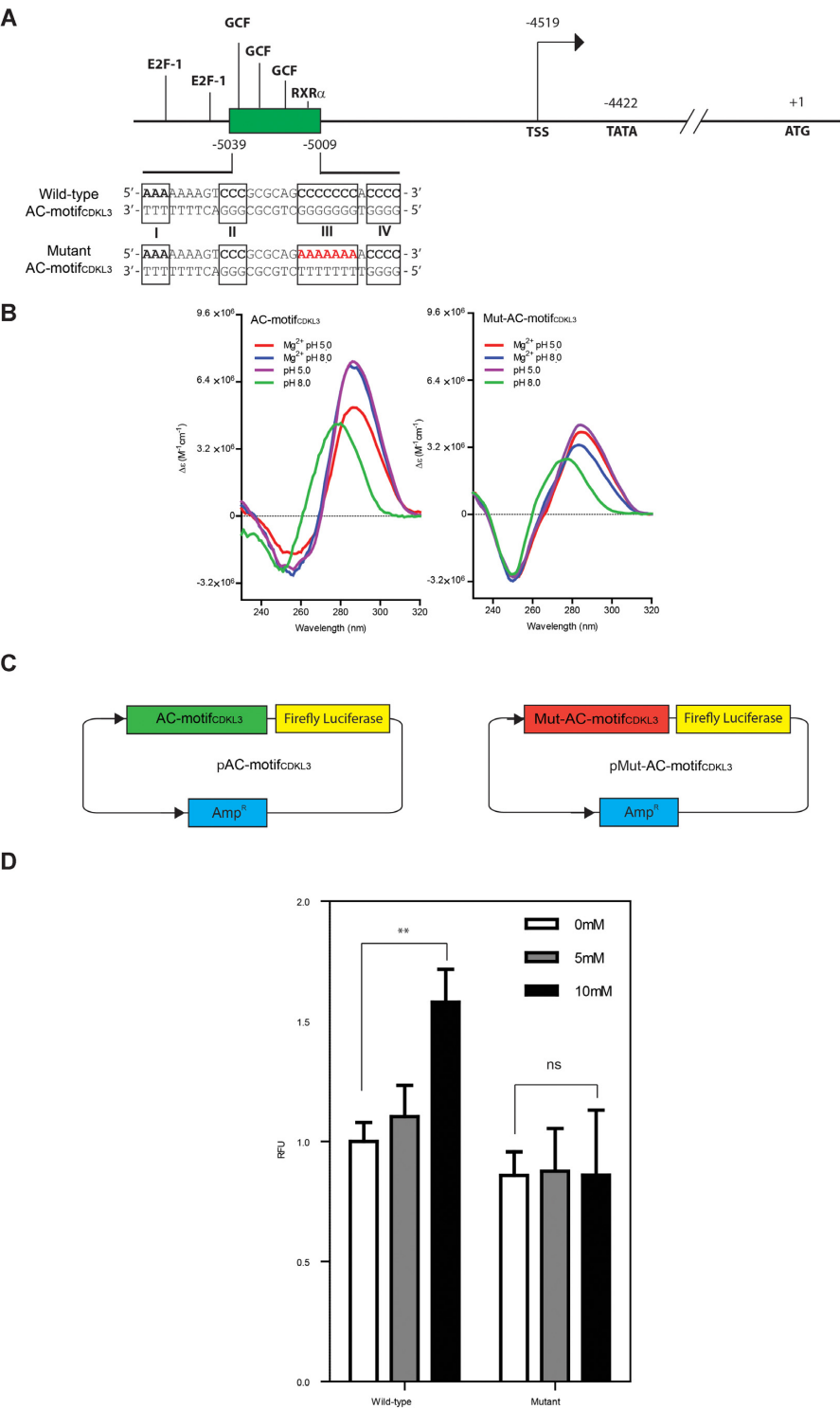


Figure 4. AC-motif formation on the CDKL3 promoter site. (A) The schematic illustrations of the AC-motif-forming site in the promoter region of CDKL3 and the sequence of AC-motif_{CDKL3}. Boxes I, II, III and IV represent the adenine and cytosine repeats participating in the formation of AC-motif structure. GCF (GC-rich sequence DNA-binding factor), RXRα (Retinoid X receptor alpha) and E2F-1 transcription factor binding sites within the CDKL3 promoter regions are indicated in bold and the sites were not altered in mutant *HeLa* cells. The start codon (ATG) is indicated at +1 position. TATA box (TATA) located at 4422 bp upstream of ATG and the transcriptional start site (TSS) at 4519 bp upstream relative to ATG are indicated. In the mutant AC-motif_{CDKL3}, the cytosine repeat in Box III is replaced by an adenine repeat. (B) The CD spectra of AC-motif_{CDKL3} and Mut-AC-motif_{CDKL3} in the presence and absence of Mg²⁺ at pH 5.0 and 8.0, respectively. (C) The schematic illustration of the pGL3-basic reporter containing 1 kb length of the wild-type (pAC-motif_{CDKL3}) and the mutant (pMut-AC-motif_{CDKL3}) promoter of CDKL3. (D) The metal-dependent luciferase activity of the reporters, pAC-motif_{CDKL3} and pMut-AC-motif_{CDKL3}. The transcriptional activity of the wild-type and the mutant CDKL3 promoter is represented by the *firefly* luciferase activity normalized to *renilla* luciferase activity. The reporter plasmids were transfected into *HEK293* cells cultured in the presence of Mg²⁺ (0, 5 and 10 mM). ***P*(0.0031) < 0.05, *n* = 3, *R*² = 9098, ns = *P*(0.9942); not significant. Statistical significance was assessed by two-tailed *t*-test.

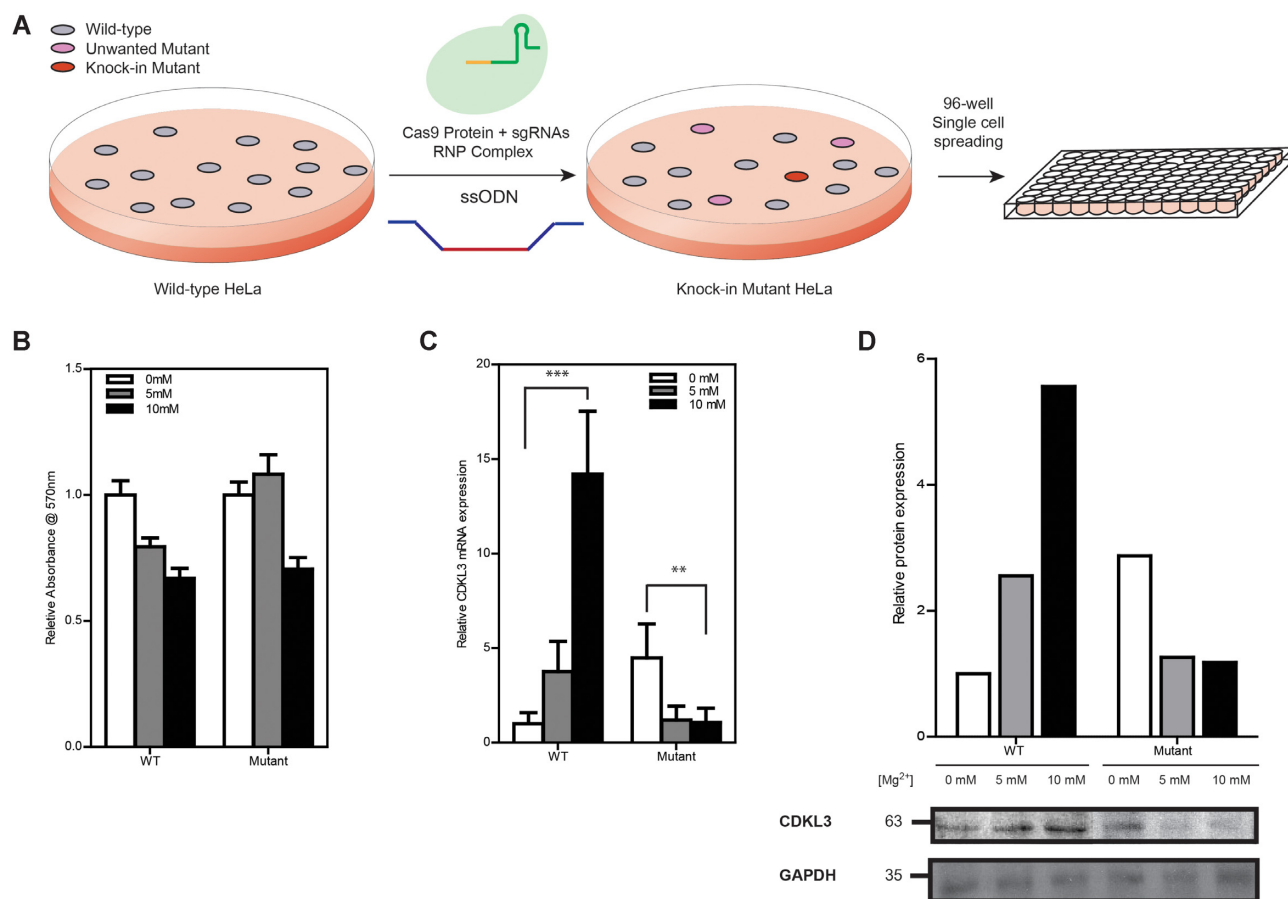


Figure 5. Magnesium-dependent expression of CDKL3 in wild-type and knock-in mutant *HeLa* cell lines. (A) Schematic representation of producing knock-in mutant *HeLa* cell lines. (B) Mg^{2+} -dependent cell viability of *HeLa* cells. Cell viability of the wild-type (WT) and the mutant *HeLa* cells was measured by MTT assay in the presence of Mg^{2+} at the concentration of 0, 5 and 10 mM. Mutant *HeLa* cells were produced using CRISPR/Cas9 genome engineering technology. (C) The effect of Mg^{2+} on the expression of *Cdkl3* in wild-type and mutant *HeLa* cells. The *Cdkl3* expression levels of the cells cultured in media containing 0, 5 and 10 mM Mg^{2+} were quantified by real time PCR (RT-PCR) with normalization to *Gapdh*. The cells were incubated in the Mg^{2+} containing media for 24 h after changing the media. The results are displayed as the relative expression level of *Cdkl3* of the wild-type *HeLa* cells incubated at 0 mM Mg^{2+} . (D) The effect of Mg^{2+} on the protein levels of CDKL3 in the wild-type and the mutant *HeLa* cells. The cells were incubated in the same conditions as Figure 5b. The protein expression of CDKL3 was quantified by band intensity of the western blots with normalization to GAPDH protein. The quantified band intensity is displayed as the relative amount to the proteins in wild-type *HeLa* cells incubated at 0 mM Mg^{2+} . ** $P(0.0082) < 0.01$, $n = 3$, $R^2 = 9289$, *** $P(0.0003) < 0.001$, $n = 3$, $R^2 = 0.9928$. Statistical significance was assessed by two-tailed *t*-test.

aspect, **AC-motif** is expected to contain a non-canonical $A^+ : C$ base pairs intercalated with $C^+ : C$ base pairs with similar secondary structure to that of canonical i-motif (Figure 6) based on the CD, fluorescence, and NMR spectra. Since pK_a values of adenine is less than cytosine, $A : C^+$ is preferred than $A^+ : C$ when they form hydrogen bond. However, we proposed that the protonation is achieved on adenine but not on cytosine in $A : C$ base pair upon **AC-motif** formation under current experimental conditions based on the following observations (Figure 2B): the cross-peaks of $[A]N1H/[A]H2$ and $[A]N1H/[A]N6H'$ that are appeared in 'Box a' (Figure 2B) would likely to be the result of the protonation at N1 of adenine in $A : C$ base pair rather than the protonation at N3 of cytosine of which its chemical shift is observed at around 15.5 ppm (5). Consistently, we also verified that C protonation is not preferred from the MD simulation result (Supplementary Figure S10). In this experiment, the structural model **AC-motif** of which cytosine is protonated, and adenine is deprotonated (**cpAC-motif**) dis-

integrated within 500 ns of molecular dynamics (MD) simulation (Supplementary Figure S10B) while the same model with adenine protonation maintains the structural integrity during the simulation (Supplementary Figure S10A) These results suggest that the protonation on adenine is more favored than cytosine in achieving a stable **AC-motif** structure. Nevertheless, the high resolution atomic structure will be necessary for further verification of the atomic interaction in the **AC-motif**.

In general, protonated adenine and cytosine do not involve in the formation of base pair in physiological pH because their pK_a (< 4.5) is lower than the neutral pH. However, it has been reported that the protonation in canonical $A^+ : C$ wobble base pair can be achieved in higher order nucleic acid structure such as riboswitch (37). This is because the protonation may not solely depend upon pK_a values of isolated nucleotide base, but rather due to the alteration of the local pK_a caused by charge redistribution among nucleotide bases upon forming a higher order

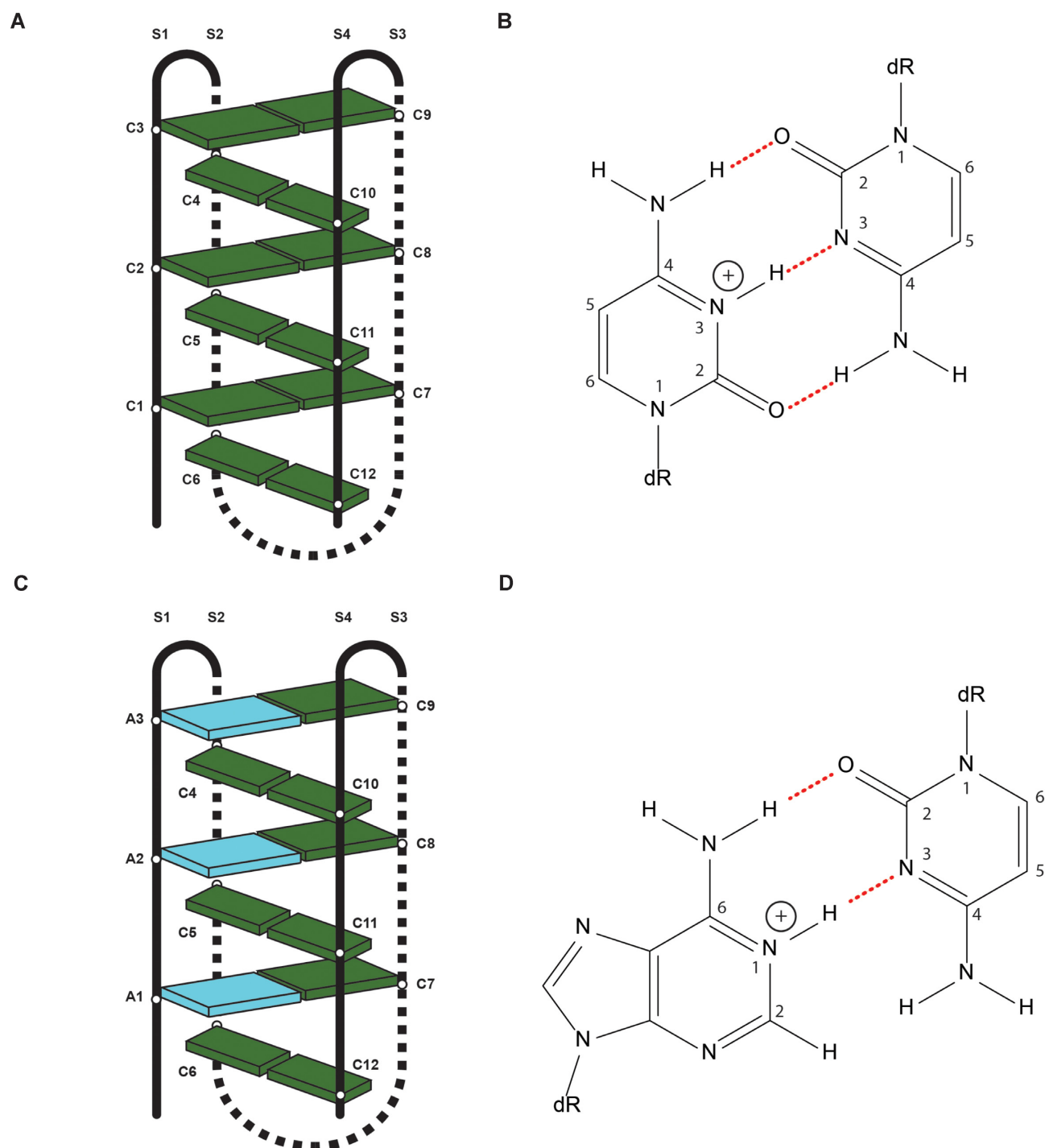


Figure 6. Schematic representation of i-motif and AC-motif. (A) Schematic representation of overall folding of i-motif. Two parallel strands (S1 and S3) bonded by non-Watson-Crick base pairings between hemi-protonated cytosine and cytosine (C⁺:C) are intercalated with other parallel strands (S2 and S4) in the opposite direction to form an intercalated structure. (B) Schematic representation of non-Watson-Crick base pairing between the hemi-protonated cytosine (C⁺) at the N3 position and cytosine in the i-motif structure. (C) Schematic representation of the proposed overall folding of AC-motif. Two parallel strands (S1 and S3) bonded by non-Watson-Crick base pairings between hemi-protonated adenine (cyan) and cytosine (green) (A⁺:C) are intercalated with other parallel strands (S2 and S4) in the opposite direction with a C⁺:C base pair to form an intercalated structure. (D) Schematic representation of the putative non-Watson-Crick base pairing between the hemi-protonated adenine (A⁺) at the N1 position and cytosine (C) in the AC-motif structure. Nucleotide sequences and strand numbers are indicated (Figure 6A and C), and atom numbers in adenine and cytosine are indicated (Figure 6B and D). Cytosine and protonated adenine bases are drawn as green and cyan plates, respectively. Possible hydrogen bonds are indicated by red dotted lines.

nucleic acid structure (37,46–48). Consistently, i-motif formation at near physiological pH is also observed *in vitro* (22) and *in vivo* (18) conditions including human cells monitored with in-cell NMR (49). Based on these evidences, we also speculate the reason behind the formation of non-canonical A⁺:C base pairings via protonation of adenine within the context of AC-motif might be due to the alteration of local pK_a that enables the protonation of adenine upon formation of higher order structure.

There was controversy over the biological implication of the i-motif structure until recent antibody detection of i-motif in the nucleus (19), since the i-motif structure is not expected to form at physiological conditions. We found that AC-motif can form a stable i-motif-like structure at physiological conditions in the presence of metal, and thus AC-motif is likely to be present in the genome and involved in many cellular events. We have proved that the many AC-motif-forming sequences in the genome actually form a secondary structure. Furthermore, we demonstrated that AC-motif formation is relevant to the regulation of the gene expression in response to Mg²⁺.

In this study, we found that magnesium ion has a critical role in stabilizing the AC-motif structure in the physiological condition. The role of magnesium ion in the formation and stabilization of i-motif was also suggested from the study where i-motif structure is induced and stabilized at physiological solution under molecular crowding condition in the presence of magnesium ion (34). It was also reported that (Ag⁺) and copper I (Cu⁺) ions stabilize the i-motif structure at neutral pH (50,51). In these cases, metal cation is proposed to stabilize structure by forming a Cytosine–Metal⁺–Cytosine structure (50–52). However, Mg²⁺ binding between two bases is not expected in AC-motif since metal-dependent change in NMR spectra was not observed (Figure 2). Although the exact role of Mg²⁺ in the formation of AC-motif cannot clearly be defined with our current data, we learned that Mg²⁺ plays an imperative role in maintaining the AC-motif structure in a relatively neutral pH environment, possibly by stabilizing the negatively charged sugar phosphate backbone or by stabilizing the long-loop structure of AC-motif (53,54). The role of Mg²⁺ ion is previously reported as for maintaining the non-canonical DNA including Holliday junction of which is stabilized in the presence of 10 mM of Mg²⁺ ion (55). Indeed, among reported 509 DNA crystals, 113 crystal structures contain magnesium ion bound either to phosphate backbones or bases (56), suggesting that magnesium ion might also contributes to the stability of AC-motif. In addition, the effect of magnesium ion on CDKL3 gene expression could be further confirmed by microarray experiments, in which fold change was 1.54 in the wild-type cell while such was 1.16 in the KI mutant cells. Interestingly, our microarray data also showed the higher responsiveness of genes containing AC-motif in their promoters to Mg²⁺ ion, compared to other genes without AC-motif in their promoter regions (Supplementary Figure S17). Further analyses of the genes containing AC-motifs will elucidate the biological significance of AC-motifs in the genome. Considering the neutral pH condition and the abundance of metals in cell, report on the existence of i-motif in nuclei (18) and our current cell biology results suggest that metal ions are important in maintaining

the stability of the DNA secondary structure and their roles in cells.

By searching DNA motifs with adenine- and cytosine-repeats connected by hexa-nucleotides as a linker, we have identified 2151 AC-motif-forming sequences in the human genome. However, considering the i-motif-like CD spectra of AC-motif with different linker size (Supplementary Figure S4), we can expect numerous AC-motif-forming sequences to be present in human DNA and RNA in the same way as G-quadruplex DNAs and RNAs (57–59). We investigated AC-motif_{CDKL3} present in the promoter region of CDKL3 as a model case and found that it up-regulated the transcription of CDKL3 dependent on Mg²⁺ concentration. However, by extrapolating the biological roles of G-quadruplex into AC-motif, and considering the position of AC-motif-forming sequence in several places in the genome such as splicing sites, introns, exons and un-translated regions (UTRs), AC-motif is expected to be involved in various biological activities. In the case of HRAS i-motif, the formation of i-motif down-regulates the transcription (60). On the other hand, our study demonstrates the up-regulation of CDKL3 gene upon formation of AC-motif in the promoter region. We speculate that the binding of some transcription factors interacting with AC-motif could increase CDKL3 gene expression, in a manner similar to ATRX protein upregulating the expression of Xlr3b gene by binding to the G-quadruplex present in the promoter region of Xlr3b gene (61). Further study should be conducted to further explore this intriguing idea. Moreover, we also think that the formation of AC-motif can dynamically compete with i-motif depending on the biological circumstances such as different transcription factors binding or negative supercoiling since they might seem to share similar cytosine tracts as AC-motif forming sequences overlap with i-motif forming sequences in many cases (Supplementary Table S1). Until now, the biological implication of many DNA repeats has not been well understood. As demonstrated in this study, their biological activities may be linked to stable non-canonical structures, which are not expected or predicted based on current knowledge. Therefore, for a comprehensive understanding of those DNA motifs with repeated sequences at the molecular and genomic level, it is necessary to conduct further studies on their structure and function. Our current study contributes to future investigation on nucleic acid secondary structures by providing a platform to identify and verify the structure and function of novel nucleic acid motifs.

DATA AVAILABILITY

Derived data supporting the findings of this study are available from the corresponding author (KK) on request. All the raw microarray data and corresponding outputs have been deposited to the Gene Expression Omnibus (GEO) database with the accession number GSE181305.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

FUNDING

Samsung Science & Technology Foundation [SSTF-BA1301-01]; National Research Foundation of Korea [2020R1A4A1018019, 2021R1A2C3011644 to K.K., 2021R1A2C3012908 to S.B.]; Korea Basic Science Institute [C140440]; Ministry of Science and ICT (to C.-J.P.). Funding for open access charge: National Research Foundation of Korea [2021R1A2C3011644].
Conflict of interest statement. None declared.

REFERENCES

- Crawford, J.L., Kolpak, F.J., Wang, A.H., Quigley, G.J., van Boom, J.H., van der Marel, G. and Rich, A. (1980) The tetramer d(CpGpCpG) crystallizes as a left-handed double helix. *Proc. Natl. Acad. Sci. USA*, **77**, 4016–4020.
- Conner, B.N., Takano, T., Tanaka, S., Itakura, K. and Dickerson, R.E. (1982) The molecular structure of d(ICpCpGpG), a fragment of right-handed double helical A-DNA. *Nature*, **295**, 294–299.
- Panayotatos, N. and Wells, R.D. (1981) Cruciform structures in supercoiled DNA. *Nature*, **289**, 466–470.
- Rhodes, D. and Giraldo, R. (1995) Telomere structure and function. *Curr. Opin. Struct. Biol.*, **5**, 311–322.
- Leroy, J.L., Gueron, M., Mergny, J.L. and Helene, C. (1994) Intramolecular folding of a fragment of the cytosine-rich strand of telomeric DNA into an i-motif. *Nucleic Acids Res.*, **22**, 1600–1606.
- Chakraborty, S., Sharma, S., Maiti, P.K. and Krishnan, Y. (2009) The poly dA helix: a new structural motif for high performance DNA-based molecular switches. *Nucleic Acids Res.*, **37**, 2810–2817.
- Paeschke, K., Bochman, M.L., Garcia, P.D., Cejka, P., Friedman, K.L., Kowalczykowski, S.C. and Zakian, V.A. (2013) Pif1 family helicases suppress genome instability at G-quadruplex motifs. *Nature*, **497**, 458–462.
- Wolff, S., Wittig, B. and Rich, A. (1995) Identification of transcriptionally induced Z-DNA segments in the human c-myc gene. *Biochim. Biophys. Acta*, **1264**, 294–302.
- Wittig, B., Wolff, S., Dorbic, T., Vahrson, W. and Rich, A. (1992) Transcription of human c-myc in permeabilized nuclei is associated with formation of Z-DNA in three discrete regions of the gene. *EMBO J.*, **11**, 4653–4663.
- Ambrus, A., Chen, D., Dai, J., Bialis, T., Jones, R.A. and Yang, D. (2006) Human telomeric sequence forms a hybrid-type intramolecular G-quadruplex structure with mixed parallel/antiparallel strands in potassium solution. *Nucleic Acids Res.*, **34**, 2723–2735.
- Sen, D. and Gilbert, W. (1988) Formation of parallel four-stranded complexes by guanine-rich motifs in DNA and its implications for meiosis. *Nature*, **334**, 364–366.
- Siddiqui-Jain, A., Grand, C.L., Bearss, D.J. and Hurley, L.H. (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.
- Kaiser, C.E., Van Ert, N.A., Agrawal, P., Chawla, R., Yang, D. and Hurley, L.H. (2017) Insight into the complexity of the i-Motif and G-Quadruplex DNA structures formed in the KRAS promoter and subsequent drug-induced gene repression. *J. Am. Chem. Soc.*, **139**, 8522–8536.
- Onel, B., Carver, M., Wu, G., Timonina, D., Kalarn, S., Larriva, M. and Yang, D. (2016) A new G-Quadruplex with hairpin loop immediately upstream of the human BCL2 P1 promoter modulates transcription. *J. Am. Chem. Soc.*, **138**, 2563–2570.
- Grand, C.L., Powell, T.J., Nagle, R.B., Bearss, D.J., Tye, D., Gleason-Guzman, M. and Hurley, L.H. (2004) Mutations in the G-quadruplex silencer element and their relationship to c-MYC overexpression, NM23 repression, and therapeutic rescue. *Proc. Natl. Acad. Sci. USA*, **101**, 6140–6145.
- Dhakal, S., Yu, Z., Konik, R., Cui, Y., Koirala, D. and Mao, H. (2012) G-quadruplex and i-motif are mutually exclusive in ILPR double-stranded DNA. *Biophys. J.*, **102**, 2575–2584.
- Kang, H.J., Kendrick, S., Hecht, S.M. and Hurley, L.H. (2014) The transcriptional complex between the BCL2 i-motif and hnRNP LL is a molecular switch for control of gene expression that can be modulated by small molecules. *J. Am. Chem. Soc.*, **136**, 4172–4185.
- Niu, K., Zhang, X., Deng, H., Wu, F., Ren, Y., Xiang, H., Zheng, S., Liu, L., Huang, L., Zeng, B. et al. (2017) BmILF and i-motif structure are involved in transcriptional regulation of BmPOUM2 in Bombyx mori. *Nucleic Acids Res.*, **46**, 1710–1723.
- Zeraati, M., Langley, D.B., Schofield, P., Moye, A.L., Rouet, R., Hughes, W.E., Bryan, T.M., Dinger, M.E. and Christ, D. (2018) I-motif DNA structures are formed in the nuclei of human cells. *Nat. Chem.*, **10**, 631–637.
- Gehring, K., Leroy, J.L. and Gueron, M. (1993) A tetrameric DNA structure with protonated cytosine-cytosine base pairs. *Nature*, **363**, 561–565.
- Sun, D. and Hurley, L.H. (2009) The importance of negative superhelicity in inducing the formation of G-quadruplex and i-motif structures in the c-Myc promoter: implications for drug targeting and control of gene expression. *J. Med. Chem.*, **52**, 2863–2874.
- Rajendran, A., Nakano, S. and Sugimoto, N. (2010) Molecular crowding of the cosolutes induces an intramolecular i-motif structure of triplet repeat DNA oligomers at neutral pH. *Chem. Commun.*, **46**, 1299–1301.
- Cantor, C.R., Warshaw, M.M. and Shapiro, H. (1970) Oligonucleotide interactions. 3. Circular dichroism studies of the conformation of deoxypolynucleotides. *Biopolymers*, **9**, 1059–1077.
- Sklenar, V. and Bax, A. (1987) Spin-echo water suppression for the generation of pure-phase two-dimensional NMR spectra. *J. Magn. Reson.*, **74**, 469–479.
- Case, D.C., Cheatham, T., Darden, T., Duke, R., Giese, T.J., Gohlke, H., Götz, A., Greene, D., Homeyer, N., Izadi, S. et al. (2017) In: *Amber 2017*. University of California, San Francisco.
- Roe, D.R. and Cheatham, T.E., 3rd (2013) PTRAJ and CPPTRAJ: Software for processing and analysis of molecular dynamics trajectory data. *J. Chem. Theory Comput.*, **9**, 3084–3095.
- Vaught, A. (1996) Graphing with Gnuplot and Xmgr: two graphing packages available under linux. *Linux J.*, **1996**, 7.
- Jorgensen, W. (1981) Transferable intermolecular potential functions for water, alcohols and ethers. Application to liquid water. *J. Am. Chem. Soc.*, **103**, 335–340.
- Ryckaert, J.P., Ciccotti, G. and Berendsen, H.J. (1977) Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *J. Comput. Phys.*, **23**, 327–341.
- Berendsen, H.J., Postma, J.V., van Gunsteren, W.F., DiNola, A.R.H.J. and Haak, J.R. (1984) Molecular dynamics with coupling to an external bath. *J. Chem. Phys.*, **81**, 3684–3690.
- Park, J., Lim, K., Kim, J.S. and Bae, S. (2017) Cas-analyzer: an online tool for assessing genome editing results using NGS data. *Bioinformatics*, **33**, 286–288.
- Tajmir-Riahi, H.A., Langlais, M. and Savoie, R. (1988) A laser Raman spectroscopic study of the interaction of calf-thymus DNA with Cu(II) and Pb(II) ions: metal ion binding and DNA conformational changes. *Nucleic Acids Res.*, **16**, 751–762.
- Langlais, M., Tajmir-Riahi, H.A. and Savoie, R. (1990) Raman spectroscopic study of the effects of Ca²⁺, Mg²⁺, Zn²⁺, and Cd²⁺ ions on calf thymus DNA: binding sites and conformational changes. *Biopolymers*, **30**, 743–752.
- Saxena, S., Joshi, S., Shankaraswamy, J., Tyagi, S. and Kukreti, S. (2017) Magnesium and molecular crowding of the cosolutes stabilize the i-motif structure at physiological pH. *Biopolymers*, **107**, e23018.
- Dai, J., Hatzakis, E., Hurley, L.H. and Yang, D. (2010) I-motif structures formed in the human c-MYC promoter are highly dynamic—insights into sequence redundancy and I-motif stability. *PLoS One*, **5**, e11647.
- Bhattacharya, P.K., Cha, J. and Barton, J.K. (2002) ¹H NMR determination of base-pair lifetimes in oligonucleotides containing single base mismatches. *Nucleic Acids Res.*, **30**, 4740–4750.
- Keller, H., Weickhmann, A.K., Bock, T. and Wohnert, J. (2018) Adenine protonation enables cyclic-di-GMP binding to cyclic-GAMP sensing riboswitches. *RNA*, **24**, 1390–1402.
- Phan, A.T. and Leroy, J.L. (2000) Intramolecular i-motif structures of telomeric DNA. *J. Biomol. Struct. Dyn.*, **17**, 245–251.
- Boulard, Y., Cognet, J.A., Gabarro-Arpa, J., Le Bret, M., Carbonnaux, C. and Fazakerley, G.V. (1995) Solution structure of an oncogenic DNA duplex, the K-ras gene and the sequence containing

- a central C.A or A.G mismatch as a function of pH: nuclear magnetic resonance and molecular dynamics studies. *J. Mol. Biol.*, **246**, 194–208.
40. Ulrich, E.L., Akutsu, H., Doreleijers, J.F., Harano, Y., Ioannidis, Y.E., Lin, J., Livny, M., Mading, S., Maziuk, D., Miller, Z. *et al.* (2008) BioMagResBank. *Nucleic Acids Res.*, **36**, D402–D408.
 41. Leroy, J.L. and Gueron, M. (1995) Solution structures of the i-motif tetramers of d(TCC), d(5methylCCT) and d(T5methylCC): novel NOE connections between amino protons and sugar protons. *Structure*, **3**, 101–120.
 42. Esmaili, N. and Leroy, J.L. (2005) i-motif solution structure and dynamics of the d(AACCCC) and d(CCCCCA) tetrahymena telomeric repeats. *Nucleic Acids Res.*, **33**, 213–224.
 43. Weil, J., Min, T., Yang, C., Wang, S., Sutherland, C., Sinha, N. and Kang, C. (1999) Stabilization of the i-motif by intramolecular adenine-adenine-thymine base triple in the structure of d(ACCCT). *Acta Crystallograph. Section D, Biol. Crystallogr.*, **55**, 422–429.
 44. Dreos, R., Ambrosini, G., Groux, R., Cavin Perier, R. and Bucher, P. (2017) The eukaryotic promoter database in its 30th year: focus on non-vertebrate organisms. *Nucleic Acids Res.*, **45**, D51–D55.
 45. Jaluria, P., Betenbaugh, M., Konstantopoulos, K. and Shiloach, J. (2007) Enhancement of cell proliferation in various mammalian cell lines by gene insertion of a cyclin-dependent kinase homolog. *BMC Biotech.*, **7**, 71.
 46. Tang, C.L., Alexov, E., Pyle, A.M. and Honig, B. (2007) Calculation of pKas in RNA: on the structural origins and functional roles of protonated nucleotides. *J. Mol. Biol.*, **366**, 1475–1496.
 47. Halder, A., Halder, S., Bhattacharyya, D. and Mitra, A. (2014) Feasibility of occurrence of different types of protonated base pairs in RNA: a quantum chemical study. *Phys. Chem. Chem. Phys.*, **16**, 18383–18396.
 48. Wolter, A.C., Weickmann, A.K., Nasiri, A.H., Hantke, K., Ohlenschläger, O., Wunderlich, C.H., Kreutz, C., Duchardt-Ferner, E. and Wohnert, J. (2017) A stably protonated adenine nucleotide with a highly shifted pKa value stabilizes the tertiary structure of a GTP-Binding RNA aptamer. *Angew. Chem.*, **56**, 401–404.
 49. Dzatko, S., Krafčíková, M., Hansel-Hertsch, R., Fessl, T., Fiala, R., Loja, T., Krafčík, D., Mergny, J.L., Foldynova-Trantírková, S. and Trantírek, L. (2018) Evaluation of the stability of DNA i-Motifs in the Nuclei of Living Mammalian Cells. *Angew. Chem.*, **57**, 2165–2169.
 50. Day, H.A., Huguin, C. and Waller, Z.A. (2013) Silver cations fold i-motif at neutral pH. *Chem. Commun.*, **49**, 7696–7698.
 51. Gao, J., Berden, G., Rodgers, M.T. and Oomens, J. (2016) Interaction of Cu(+) with cytosine and formation of i-motif-like C-M(+)-C complexes: alkali versus coinage metals. *Phys. Chem. Chem. Phys.*, **18**, 7269–7277.
 52. Abdelhamid, M.A., Fabian, L., MacDonald, C.J., Cheesman, M.R., Gates, A.J. and Waller, Z.A. (2018) Redox-dependent control of i-Motif DNA structure using copper cations. *Nucleic Acids Res.*, **46**, 5886–5893.
 53. Lipfert, J., Doniach, S., Das, R. and Herschlag, D. (2014) Understanding nucleic acid-ion interactions. *Annu. Rev. Biochem.*, **83**, 813–841.
 54. Takahashi, S. and Sugimoto, N. (2020) Stability prediction of canonical and non-canonical structures of nucleic acids in various molecular environments and cells. *Chem. Soc. Rev.*, **49**, 8439–8468.
 55. Zettl, T., Shi, X., Bonilla, S., Sedlak, S.M., Lipfert, J. and Herschlag, D. (2020) The structural ensemble of a Holliday junction determined by X-ray scattering interference. *Nucleic Acids Res.*, **48**, 8090–8098.
 56. Shamim, A., Parveen, N., Subramani, V.K. and Kim, K.K. (2020) Molecular packing interaction in DNA crystals. *Crystals*, **19**, 1093.
 57. Tippiana, R., Xiao, W. and Myong, S. (2014) G-quadruplex conformation and dynamics are determined by loop length and sequence. *Nucleic Acids Res.*, **42**, 8106–8114.
 58. Pandey, S., Agarwala, P. and Maiti, S. (2013) Effect of loops and G-quartets on the stability of RNA G-quadruplexes. *J. Phys. Chem. B*, **117**, 6896–6905.
 59. Zhang, A.Y., Bugaut, A. and Balasubramanian, S. (2011) A sequence-independent analysis of the loop length dependence of intramolecular RNA G-quadruplex stability and topology. *Biochemistry*, **50**, 7251–7258.
 60. Miglietta, G., Cogoi, S., Pedersen, E.B. and Xodo, L.E. (2015) GC-elements controlling HRAS transcription form i-motif structures unfolded by heterogeneous ribonucleoprotein particle A1. *Sci. Rep.*, **5**, 18097.
 61. Shioda, N., Yabuki, Y., Yamaguchi, K., Onozato, M., Li, Y., Kurosawa, K., Tanabe, H., Okamoto, N., Era, T., Sugiyama, H. *et al.* (2018) Targeting G-quadruplex DNA as cognitive function therapy for ATR-X syndrome. *Nat. Med.*, **24**, 802–813.