

# Effect of ATRX and G-quadruplex formation by the VNTR sequence on $\alpha$ -globin gene expression

Yue Li,<sup>[a]</sup> Junetha Syed,<sup>[a]</sup> Yuki Suzuki,<sup>[a]</sup> Sefan Asamitsu,<sup>[a]</sup> Norifumi Shioda,<sup>[c]</sup> Takahito Wada,<sup>[d]</sup> and Hiroshi Sugiyama,<sup>\*,[a, b]</sup>

**Abstract:** ATR-X ( $\alpha$ -thalassemia/mental retardation syndrome X-linked) syndrome is caused by mutations in chromatin-remodeler ATRX. ATRX can bind the variable number of tandem repeats (VNTR) sequence in the promoter region of the  $\alpha$ -globin gene cluster. The VNTR sequence, which contains the potential G-quadruplex-forming sequence CGC(GGGGCGGG)<sub>n</sub>, is involved in the downregulation of  $\alpha$ -globin expression. We investigated the G-quadruplex and i-motif formation in the single-stranded DNA and long double-stranded DNA. The promoter region without the VNTR sequence showed approximately two fold higher luciferase activity than the promoter region harboring VNTR sequence. G-quadruplex stabilizers hemin and TMPyP4 reduced the luciferase activity whereas expression of ATRX led to a recovery in reporter activity. Our results demonstrate that stable G-quadruplex formation by the VNTR sequence downregulates the expression of  $\alpha$ -globin genes and that ATRX might bind to and resolve the G-quadruplex.

## Introduction

ATRX is a SWI/SNF-like chromatin remodeler that can bind to the histone chaperone DAXX and incorporate the variant histone H3.3 into telomeric and pericentromeric heterochromatin.<sup>[1-5]</sup> It comprises several conserved protein-binding domains that are critical to its functions, such as the ATRX-DNMT3-DNMT3L (ADD) domain and the helicase domain.<sup>[6, 7]</sup> The ADD domain was found to bind histone H3 by recognizing lysine 9 trimethylation (H3K9me3)<sup>[8, 9]</sup> and the ATP-dependent triplex displacement activity of ATRX-DAXX complex suggested the potential function of helicase domain in the presence of ATP.<sup>[1]</sup> Numerous mutations located within the ATRX protein coding sequence give rise to ATR-X syndrome, which is characterized by developmental disabilities,  $\alpha$ -thalassemia, and cognitive deficits.<sup>[10-14]</sup> Studies on  $\alpha$ -thalassemia suggested that ATRX could affect  $\alpha$ -globin expression by controlling the incorporation of repressive macroH2A1 histone variant to the chromatin.<sup>[15]</sup> Mouse models of the ATR-X mental retardation syndrome are characterized by the activation of imprinted genes and abnormal

dendritic spine formation<sup>[16-18]</sup> due to the loss of functional ATRX in the neonatal brain. However, the underlying mechanism of how these ATRX mutations induce the symptoms of ATR-X syndrome is not clear and effective treatments for ATR-X syndrome are yet to be developed. Recent studies also demonstrated the role of functional ATRX as tumor suppressor protein that negatively regulates the process of alternative lengthening of telomeres (ALT) in cancer cells.<sup>[19-20]</sup>

Many human disorders are caused by expandable DNA repeats, which are thought to form unusual structures that can disrupt cellular replication, repair, and recombination machineries.<sup>[21]</sup> Such repeats can be located in either coding or noncoding regions of important genes. Many of these repeats are GC-rich sequences with the potential to form G-quadruplex structures, which are stabilized by Hoogsteen hydrogen bonds and central cations such as K<sup>+</sup> or Na<sup>+</sup>.<sup>[22-24]</sup> For example, in the coding region of *C9orf72*, a hexanucleotide repeat expansion (GGGGCC)<sub>n</sub> forms DNA G-quadruplexes that stalls the transcription machinery and produces G-quadruplexes containing truncated RNA transcripts which in turn is responsible to the pathogenesis associated with amyotrophic lateral sclerosis and frontotemporal dementia (ALS/FTD).<sup>[25, 26]</sup> Similar studies in the promoter region of human insulin gene demonstrated the formation of multiple G-quadruplexes by the insulin-linked polymorphic region (ILPR) that may regulate insulin expression in insulin-dependent diabetes mellitus.<sup>[27, 28]</sup> These results suggested that G-quadruplexes formed by GC-rich repeats might provide the basis for a mechanistic model for repeat-associated diseases. Regarding the G-quadruplexes, various forms have been identified including DNA, RNA, and DNA/RNA hybrid G-quadruplexes and the formations of G-quadruplexes have been well studied by different methods *in vitro* and *in vivo*.<sup>[29-32]</sup> Proteins with binding affinity towards these G-quadruplexes have also been identified, suggesting the importance of G-quadruplex structures in cellular processes.<sup>[33-35]</sup> Furthermore, several small molecules have been developed to target G-quadruplex formation to treat a number of diseases.<sup>[36-38]</sup>

Recently, ChIP-sequencing and gel shift studies revealed that ATRX could bind to G-quadruplex structures formed by GC-rich sequences in human genomic DNA.<sup>[39]</sup> It was proposed that ATRX could enable the passage of transcriptional machinery through GC-rich coding regions by binding to and resolving G-quadruplex structures along with the incorporation of the variant H3.3 within the transcribed genes.<sup>[39, 40]</sup> In human  $\alpha$ -globin gene clusters, a variable number of tandem repeats (VNTR) CGC(GGGGCGGG)<sub>n</sub> lies 1 kb upstream of the *HBM* promoter. The size of the VNTR is variable among individuals and it is closely related to the downregulation of  $\alpha$ -globin gene expression.<sup>[39]</sup> It seems that ATRX binds to and resolves G-quadruplex formation by the VNTR and facilitates  $\alpha$ -globin expression. However, questions concerning the G-quadruplexes

[a] Y. Li, S. Asamitsu, Dr. J. Syed, Dr. Y. Suzuki, Prof. H. Sugiyama  
Department of Chemistry, Graduate School of Science, Kyoto University, Kyoto 606-8502, Japan  
E-mail: hs@kuchem.kyoto-u.ac.jp

[b] Prof. H. Sugiyama  
Institute for Integrated Cell-Material Sciences (WPI-iCeMS), Kyoto University, Kyoto 606-8501, Japan

[c] Dr. N. Shioda  
Department of Pharmacology, Graduate School of Pharmaceutical Sciences, Tohoku University, Sendai 980-8578, Japan

[d] Prof. T. Wada  
Department of Medical Ethics and Medical Genetics, Graduate School of Medicine, Kyoto University, Kyoto 606-8501, Japan

Supporting information for this article is given via a link at the end of the document. ((Please delete this text if not appropriate))

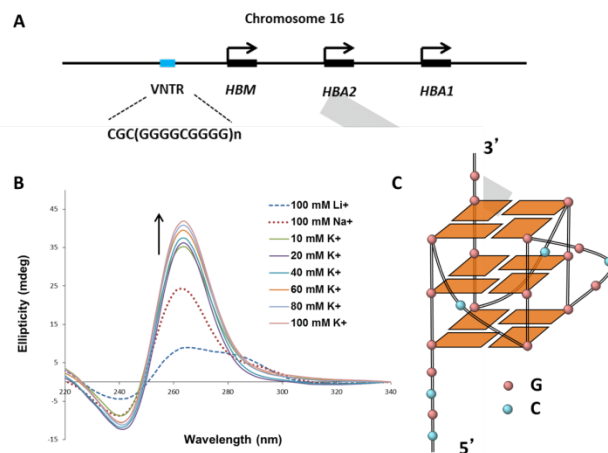
formation in VNTR sequence, the role of ATRX and how mutations in ATRX protein induce  $\alpha$ -thalassemia remain to be resolved.

In this study, we investigated the formation of G-quadruplex by the VNTR sequence using CD, UV and gel analysis. The ability of G-quadruplex formation by the VNTR sequence in long double-stranded DNA was also analysed and observed by Atomic Force Microscopy (AFM). Well-characterized G-quadruplex binding ligands such as hemin and TMPyP4 were studied for their binding affinity to the proposed G-quadruplexes formed by VNTR sequences. The gene regulatory effect of the G-quadruplex-forming VNTR sequence and the corresponding mode of action of G-quadruplex ligands and ATRX on the G-quadruplex were studied using reporter assay. Based on these results, we propose that G-quadruplex formation and ATRX play important roles in regulating the expression of  $\alpha$ -globin gene cluster. Our results also provide insights into the development of drugs for the treatment of  $\alpha$ -thalassemia induced by ATRX mutations.

## Results and Discussion

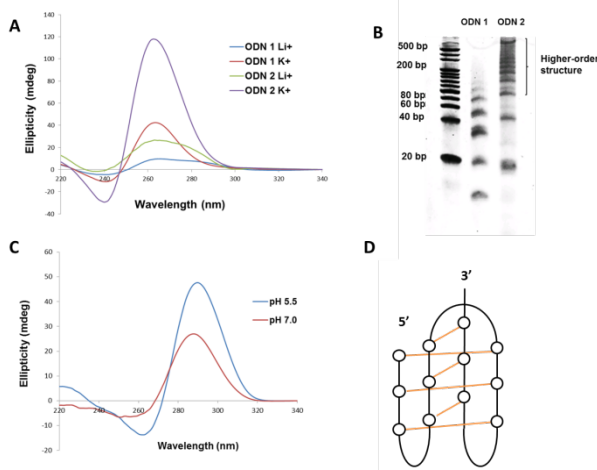
Human  $\alpha$ -globin gene expression is regulated by many factors including *cis*-acting and *trans*-acting elements and this is further complicated when the packaging of DNA into chromatin is considered.<sup>[41]</sup> Various upstream regulatory sequences in the  $\alpha$ -globin gene cluster have been identified and it was suggested that  $\alpha$ -globin gene expression could be controlled by the chromatin loops formed by upstream sequences. Recent studies indicated that VNTR sequence in the  $\alpha$ -globin gene promoter region was closely related to gene expression downstream.<sup>[39]</sup> VNTR sequence was located in the promoter region of the *HBM* gene (Figure 1A). This gene was originally found to be a pseudogene, but recent studies suggested that it also had a protein coding function.<sup>[42]</sup>

Generally, this VNTR sequence consists of CGC(GGGGCGGG)<sub>n</sub> repeats despite SNPs (single nucleotide polymorphism). The size of VNTR sequence varies among individuals and it contains a repeat sequence of approximately 360 bp in our case (Figure S1). For convenience, ODN 1 (oligodeoxynucleotide 1) with four CGGGG repeats was used to study G-quadruplex formation by the VNTR sequence. As shown in Figure 1B, CD spectra of ODN 1 showed that it could form a highly stable parallel G-quadruplex in the presence of a low concentration of potassium ions (10 mM). UV melting experiments also revealed that the parallel G-quadruplex formed by ODN 1 had high stability even in the presence of 5 mM K<sup>+</sup> (Figure S2). The sequence of ODN 1 closely resembled the G-quadruplex-forming sequence in the *RET* promoter reported previously.<sup>[43]</sup> Therefore, it may form similar parallel G-quadruplex in K<sup>+</sup> solutions; the G-quadruplex formed by ODN 1 is shown in Figure 1C.



**Figure 1.** G-quadruplex formation in the promoter region of  $\alpha$ -globin gene cluster. A) The VNTR sequence consists of CGC(GGGGCGGG)<sub>n</sub> repeats located in the *HBM* gene promoter. B) CD spectra of ODN 1 (5'-CGCGGGGCGGGGCGGGGCGGGG-3') selected from the VNTR sequence. ODN 1 formed stable parallel G-quadruplex in the presence of K<sup>+</sup>. C) Representation of the parallel G-quadruplex formed by ODN 1.

ODN 2, with eight CGGGG repeats, was used to study G-quadruplex formation by a long repeat sequence. The strong CD signals of ODN 2 compared with that of ODN 1 suggested the stacking of parallel G-quadruplexes in solutions (Figure 2A). G-quadruplex formations of ODN 1 and ODN 2 in the presence of 10 mM K<sup>+</sup> were analyzed by native gel electrophoresis. The results demonstrated that both ODN 1 and ODN 2 formed higher-order G-quadruplex structures in solutions (Figure 2B). These results are consistent with previous studies that parallel G-quadruplexes formed by DNA and RNA sequences can stack together to form higher-order structures<sup>[44, 45]</sup>. The percentage of higher-order structures formed by ODN 2 was high, whereas ODN 1 mainly formed a single G-quadruplex. Mutation studies of ODN 1 and ODN 2 were also performed as shown in Figure S2. Mutations in the G of ODN 1 or in the linker region of ODN 2 decreased the parallel G-quadruplex formation and higher-order structures, suggesting that these repeated sequences are quite important to the formation of parallel G-quadruplex and higher-order G-quadruplexes by VNTR sequence. Therefore, it is likely that the long CGC(GGGGCGGG)<sub>n</sub> repeat sequence forms complex G-quadruplex structures. The i-motif formation by the complementary sequence of ODN 1 (ODN 3) was also investigated by CD spectroscopy. As shown in Figure 2C, under acidic conditions (pH 5.5), ODN 3 showed a positive peak at about 288 nm and a negative peak at about 264 nm, indicating the formation of the i-motif structure. Together, these results suggested that the VNTR sequence with long CGC(GGGGCGGG)<sub>n</sub> repeats could form G-quadruplex and i-motif structures.



**Figure 2.** G-quadruplex and i-motif formation by the long VNTR sequence. A) CD spectra of ODN 2 with eight repeats in 100 mM  $K^+$  solution. B) G-quadruplex formation of ODN 1 and ODN 2 analysed by gel electrophoresis. C) CD spectra of the i-motif formed by the complementary sequence ODN 3 (5'-CCCCGCCCCGCCCCGCCCCGCG-3') at pH 5.5. D) Representation of the i-motif formation by the VNTR sequence.

To investigate G-quadruplex and i-motif formation in the long double-stranded DNA, we amplified two double-stranded DNA fragments of the *HBM* promoter from human genome DNA. One was approximately 0.9 kb without the VNTR sequence and a second was approximately 1.4 kb with the VNTR sequence. The G-quadruplex formation in long double-stranded DNA was observed with an AFM (Figure 3 and Figure S5). Under acidic conditions and in the presence of 100 mM KCl, bunched structures were observed in one end of the 1.4 kb DNA fragments, suggesting that the duplex DNA changed into single-stranded structures and formed G-quadruplex as well as i-motif structures. In contrast, such structures were not formed in the absence of  $K^+$  or by the 0.9 kb fragments without VNTR sequence (Figure S6). The long VNTR repeat sequence could naturally form G-quadruplex and i-motif structures in acidic condition with  $K^+$  even without annealing or *in vitro* transcription<sup>[46, 47]</sup>. In our case, the VNTR sequence is quite close to one end (Figure S4). The end of duplex could unfold and form G-quadruplex and i-motif structures with less restrictions. Therefore, we could observe branched structures instead of bright spots previously reported.<sup>[46-48]</sup> It suggested the possibility of G-quadruplex and i-motif formation during transcription or

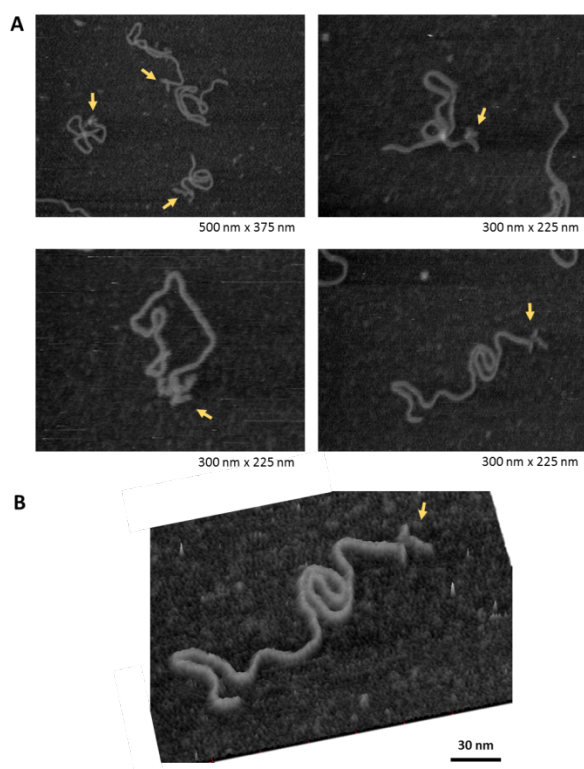
replication when the double-stranded DNA is unfolded into single-stranded DNA. Thus, it is possible that such large scale G-quadruplex formation will affect gene expressions of *HBM* and other downstream  $\alpha$ -globin genes.

It has been comprehensively demonstrated that small molecules with G-quadruplex binding affinity can be used to regulate gene expressions.<sup>[37]</sup> TMPyP4 and hemin were used to stabilize G-quadruplex formation by the VNTR sequence. TMPyP4 has high binding affinity towards G-quadruplex and hemin has been frequently used for DNazymes with G-quadruplex.<sup>[49-52]</sup> UV melting experiments were performed to analyze the stabilization effect of hemin and TMPyP4 on G-quadruplex formed by ODN 1. The  $\Delta T_m$  for hemin was 9 °C and the  $\Delta T_m$  for TMPyP4 was higher than 22 °C (Figure S3). Both TMPyP4 and hemin could stabilize G-quadruplex formation and the former showed a relatively high stabilization effect on parallel G-quadruplex formed by the VNTR sequence.

SPR experiments and the DNzyme experiments were performed to further investigate the interaction of G-quadruplex with TMPyP4 and hemin. SPR results showed that TMPyP4 had a higher binding affinity towards the parallel G-quadruplex formed by ODN 1 ( $K_D$   $6.1 \times 10^{-7}$  M) compared with hemin ( $K_D$   $2.3 \times 10^{-5}$  M) (Figure 4A and Table S1). DNzyme experiments indicated that ODN 1 exhibited a higher catalytic activity similar to that of the c-myc sequence (Figure 4B), indicating high DNzyme activity of the parallel G-quadruplex.<sup>[53]</sup> Interestingly, ODN 2 showed relatively low activity. From the gel electrophoresis results, it is possible that the high proportion of compact, higher-order structures formed by ODN 2 prohibited the binding of hemin.<sup>[44, 51]</sup>

To investigate the effect of G-quadruplex formation on  $\alpha$ -globin gene expression and the regulatory functions of small molecules and ATRX, we performed luciferase reporter assays. Two plasmids containing the 1.4 kb (with VNTR sequence) and 0.9 kb (without VNTR sequence) fragments were constructed and transfected into HEK293T cells. We then assessed the gene expression by measuring luciferase activities. For hemin and TMPyP4, different concentrations of compounds were added into medium after transfection. In the case of ATRX, halo-tagged ATRX plasmid was cotransfected with constructed plasmids to study the effect of ATRX expression on luciferase activities. The expression of ATRX in HEK293T cells was confirmed by staining with halo tag-specific ligand (Figure S7).



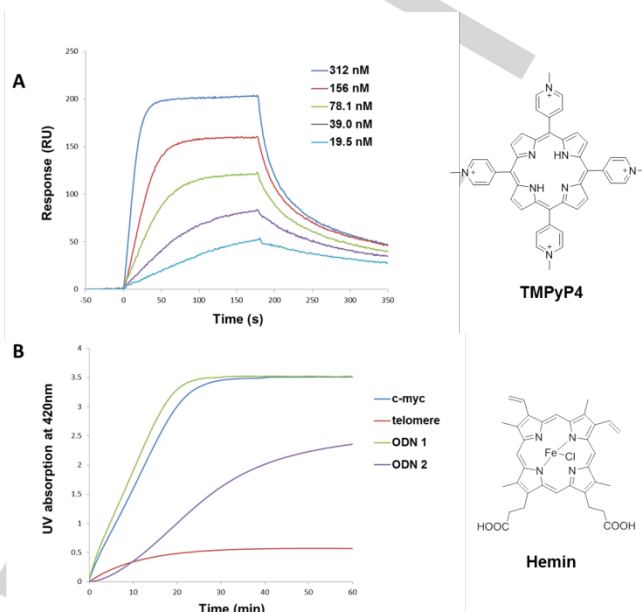


**Figure 3.** AFM images of G-quadruplex and i-motif formation by the long VNTR sequence. A) AFM images of samples of the 1.4 kb *HBM* promoter sequence at pH 5.5 in 100 mM KCl solution. B) A topographic zoomed AFM image of the DNA containing a branched structure. Positions of branched structures are indicated by yellow arrows.

As shown in Figure 5A, the luciferase activity of the 1.4 kb plasmid was lower than that of the 0.9 kb plasmid. This may be due to the downregulation effect of the formation of various G-quadruplex structures by the VNTR sequence. Since the VNTR sequence tended to form stable parallel G-quadruplexes and even higher-order structures, it might reduce the luciferase activity more effectively compared with other G-quadruplex-forming sequences reported previously.<sup>[49]</sup> These results hence suggested that G-quadruplex formation by the VNTR sequence might play an important role in  $\alpha$ -globin gene expression, in particular in the expression of the *HBM* gene.

The effect of G-quadruplex ligands were then evaluated (Figure 5B). Both hemin and TMPyP4 reduced the luciferase activities of the 1.4 kb plasmid in a concentration-dependent manner. By contrast, the luciferase activities of the 0.9 kb plasmid were slightly affected by hemin and TMPyP4. These results showed that small molecules that can stabilize parallel G-quadruplex formed by the VNTR sequence in the promoter region can downregulate gene expressions. The downregulation effect also corresponded to the stabilization effect of the G-quadruplex. TMPyP4, which had higher stabilization effect on G-quadruplex than hemin, showed a more pronounced downregulation effect on luciferase gene expression. Moreover, the downregulation

effect of TMPyP4 in current study was stronger than previous results, suggesting the effect of large scale G-quadruplex formation by VNTR sequence<sup>[47]</sup>.

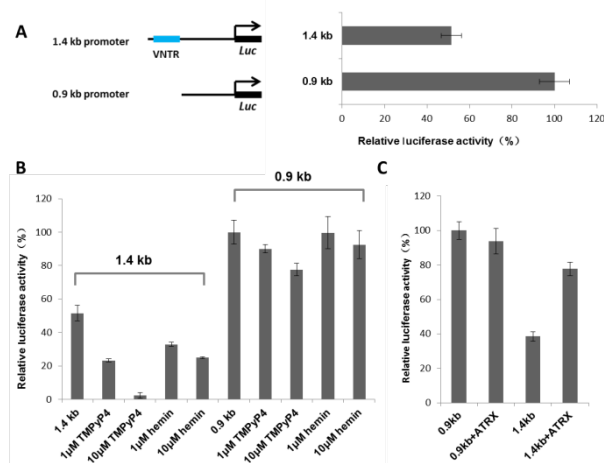


**Figure 4.** Interaction of TMPyP4 and hemin with the parallel G-quadruplex formed by ODN 1. A) SPR sensorgrams of TMPyP4 binding with ODN 1. Two-site binding model was used to fit the sensorgrams.  $K_D$  ( $6.1 \times 10^{-7}$  M) was calculated by  $K_{D1}$  ( $2.3 \times 10^{-7}$  M) and  $K_{D2}$  ( $1.6 \times 10^{-8}$  M). B) DNAzymes activity of different G-quadruplex-forming sequences including c-myc, telomere, ODN 1, and ODN 2.

Considering the binding affinity of ATRX to G-quadruplexes, it was suggested that ATRX, which has chromatin-remodeling activity, might resolve G-quadruplex formation. However, the underlying mechanism was not clear as ATRX itself failed to unfold the G-quadruplex *in vitro*.<sup>[54]</sup> Accordingly, the effect of ATRX expression on luciferase activity was analyzed. The results showed that ATRX could increase luciferase activity of the 1.4 kb plasmid (Figure 5C), suggesting that ATRX might reduce the downregulation effect of G-quadruplex formation by the VNTR sequence. These results supported the hypothesis that ATRX could resolve the potential G-quadruplex formation in GC-rich regions. The results of luciferase reporter assays shown above indicated that G-quadruplex formation by the VNTR sequence in the *HBM* promoter might have important regulatory functions on  $\alpha$ -globin gene expressions. More importantly, ATRX that could bind to GC-rich regions might be involved in and help facilitate gene expression downstream.

Human hereditary disorders that are caused by one simple repeat expansion suggested the negative impact of the non-B-form DNA structures formed by repeat sequences.<sup>[21]</sup> For a particular gene, such repeat sequences located in coding or noncoding regions would strongly disrupt normal gene expression and even generate toxic repeat RNA transcripts.<sup>[25, 26]</sup> As to ATR-X syndrome,  $\alpha$ -thalassemia, which is the downregulation of  $\alpha$ -globin expression, is related to the VNTR sequences found in the promoter region of the *HBM* gene. In the

current study, we established that stable G-quadruplex and i-motif structures could be formed by this repeat sequence and showed that more complex structural features were formed in the long double-stranded DNA. The results of luciferase experiments revealed a clear downregulation effect of the VNTR sequence on luciferase activity. In a simple model, the formation of non-B-form DNA structures would generate structural disturbances in the B-form DNA on a large scale, which could disrupt the binding of transcription factors and ultimately hamper the transcription. Given the complex regulation network in the  $\alpha$ -globin cluster, the formation of non-B-form DNA structures would affect the gene expression of other  $\alpha$ -globin genes downstream. VNTR sequence could also form G-quadruplex structures during DNA replication. These structures could stall replication forks and cause genomic and epigenetic instability. For example, reduced  $\alpha$ -globin expression could be caused by the loss of activating histone marks at the  $\alpha$ -globin gene during replication.<sup>[55, 56]</sup> Therefore, the size of the VNTR sequence is related to the downregulation effect of gene expression downstream, which is consistent with previous reports.<sup>[39]</sup>



**Figure 5.** Luciferase reporter assays of the 1.4 kb and 0.9 kb *HBM* promoter regions. A) Luciferase activity of 1.4 kb promoter sequence and 0.9 kb promoter sequence. Luciferase activity of the 1.4 kb *HBM* promoter sequence was almost two fold higher than that of the 0.9 kb *HBM* promoter sequence. B) Effect of hemin and TMPyP4 on the luciferase activity. In the presence of G-quadruplex ligands hemin and TMPyP4, luciferase activity of the 1.4 kb *HBM* promoter sequence decreased in a concentration-dependent manner, whereas luciferase activity of the 0.9 kb *HBM* promoter sequence was slightly reduced. C) Effect of ATRX expression on the luciferase activity. ATRX expression by transfection of ATRX plasmids recovered the luciferase activity of the 1.4 kb *HBM* promoter sequence. The data are mean  $\pm$  SD of repeated experiments ( $n = 3$ ).

Considering the large effect of the non-B-form DNA structures formed by repeat sequences, the resolution of these structures is important for the maintenance of gene expression in normal cells. Various helicases that can unfold G-quadruplex structures in the presence of ATP have been reported.<sup>[57, 58]</sup> Proteins that

could resolve G-quadruplex structures by shifting the duplex–quadruplex equilibrium have also been identified.<sup>[33, 57]</sup> ATRX is thought to resolve G-quadruplex based on its location on a GC-rich region and its chromatin-remodeling function. The results of our reporter assay also suggested that ATRX expression could regenerate the downregulation effect caused by the VNTR sequence. Although ATRX has the helicase domain, previous results have shown that it can bind to G-quadruplex, but cannot unfold the G-quadruplex in the presence of ATP like BLM helicase.<sup>[54]</sup> These results suggested that ATRX may resolve G-quadruplex formation by other mechanisms. Recent reports have shown that the helicase domain of ATRX could also bind to duplex DNA.<sup>[17]</sup> Therefore, ATRX may resolve G-quadruplex structures by changing the duplex–quadruplex equilibrium. It has also been shown by NMR spectroscopic and crystallographic analyses that the ADD domain of ATRX can specifically recognize a histone tag with H3K9me3.<sup>[8, 9]</sup> ATRX and DAXX together can facilitate the incorporation of histone H3.3 variant.<sup>[1]</sup> Considering these functions of ATRX, it is possible that the resolution of G-quadruplex by ATRX is related to its chromatin-remodeling activity. Overall, ATRX may unfold the G-quadruplex structure or change the duplex–quadruplex equilibrium at a nucleosome level.

For the  $\alpha$ -thalassemia in ATR-X patients, mutations are usually found in either the ADD domain or the helicase domain. These mutations would disrupt the chromatin-remodeling activity or G-quadruplex resolution function, can promote the G-quadruplex formation by the VNTR sequence, which in turn results in the downregulation of  $\alpha$ -globin gene expressions. The size of the VNTR sequences in patients is also important because it may affect the stability or scale of G-quadruplex formation. Therefore, we propose that ATRX mutations and the VNTR sequences are both involved in the pathogenesis of  $\alpha$ -thalassemia in ATR-X patients, and like other repeat expansions, the VNTR sequence tends to form non-B-form DNA during replication and transcription. Possible higher-order G-quadruplex structure may also be involved. Normally, ATRX will locate to this region, resolve G-quadruplex formation and incorporate histone 3.3 with DAXX. However, mutations in ATRX will disrupt either the location of ATRX or the resolution of the G-quadruplex by ATRX. Therefore, the G-quadruplex formed by the VNTR sequence will remain, leading to reduced levels of  $\alpha$ -globin gene expression and finally causing  $\alpha$ -thalassemia.

## Conclusions

In summary, we have investigated G-quadruplex and i-motif formation by the VNTR sequence in the promoter region of  $\alpha$ -globin gene cluster. Our findings showed that stable G-quadruplex structures could be formed by single-stranded DNA and higher-order G-quadruplex structures could be formed by long double-stranded DNA containing the VNTR sequence. The effect of G-quadruplex formation and ATRX expression were evaluated by conducting luciferase assays. The G-quadruplex formation by the VNTR sequence could downregulate luciferase

activity, whereas ATRX could reverse this downregulation effect. Our results suggested the regulatory functions of ATRX and G-quadruplex formation by VNTR in the  $\alpha$ -globin gene expressions. A model of ATRX and G-quadruplex formation was proposed and it might explain the vital disease-causing role of different size of VNTR sequence and ATRX mutations found in ATR-X patients. These results help to explain the mechanism of  $\alpha$ -thalassemia in ATR-X syndrome and provide insights into the treatment for this debilitating condition.

## Experimental Section

**CD spectra:** All oligonucleotides were purchased from Sigma-Aldrich Co. LLC. (Japan) without further purification. Oligonucleotides for CD spectra were prepared in  $\text{Li}^+$  solution (20 mM lithium cacodylate, pH 7.0, 100 mM LiCl),  $\text{Na}^+$  solution (20 mM lithium cacodylate, pH 7.0, 100 mM NaCl) or  $\text{K}^+$  solution (20 mM lithium cacodylate, pH 7.0, 100 mM KCl). Annealing was performed by heating to 95°C for 5 min and cooling down to room temperature slowly. CD spectra were measured in 0.5-nm steps from 340 to 220 nm using JASCO J-805LST Spectrometer in a 1-cm quartz cuvette.

**UV melting experiment:** UV-Vis melting temperature analyses were performed on a spectrophotometer V-650 (JASCO, Japan) with a thermocontrolled PAC-743R cell changer (JASCO, Japan) and a refrigerated and heating circulator F25-ED (Julabo). Hemin (51280) and TMPyP4 (323497) were purchased from Sigma-Aldrich Co. LLC. (Japan). Oligonucleotides (5  $\mu\text{M}$ ) in  $\text{K}^+$  solution (20 mM lithium cacodylate, pH 7.0, 5 mM KCl) and 1 eq compounds were added. Annealing was performed by heating to 95°C for 5 min and cooling down to room temperature slowly. Melting temperature ( $T_m$ ) of each sample were measured at 295 nm from 20 to 90°C at a rate of 1.0 °C /min.  $\Delta T_m$  were calculated by the  $T_m$  values of samples with and without compounds.

**Gel electrophoresis:** For the single-stranded oligonucleotides, DNA samples for gel electrophoresis were prepared similarly with the samples of CD spectra while the KCl concentration was 10 mM. Oligonucleotides were prepared in  $\text{K}^+$  solution (20 mM lithium cacodylate, pH 7.0, 10 mM KCl). Annealing was performed by heating to 95°C for 5 min and cooling down to room temperature slowly. Each sample was analysed by native gel (16%) in 1×TBE buffer containing 10 mM KCl at room temperature. The single-stranded DNA was stained with SYBR® Gold nucleic acid gel stain (S-11494, Thermo Fisher Scientific Inc. Japan). 20 bp DNA Ladder (3420A, Takara, Japan) was used as marker. The gels were then imaged by FLA-3000 (FUJIFILM, Japan).

**AFM imaging:** The 1.4 kb and 0.9 kb HBM promoter sequences were amplified from human genomic DNA using AccuPrime™ GC-Rich DNA Polymerase (Life Technologies, Japan). The DNA fragments were diluted to a concentration of 0.5 ng/ $\mu\text{L}$  in a buffer containing 20 mM MES-NaOH (pH 5.5) and 100 mM KCl. After incubation at room temperature for 1 hour, 2  $\mu\text{L}$  of the samples were deposited onto freshly cleaved mica discs pretreated with 0.1% 3-aminopropyltriethoxy silane (APTES). After 3 min incubation, the sample was rinsed with 10  $\mu\text{L}$  washes of the buffer, and imaged in the same buffer. AFM imaging was performed using a high-speed AFM system (Nano Live Vision, RIBM, Tsukuba, Japan) and small cantilevers with the dimensions ( $L \times W \times H$ ) of  $10 \times 2 \times 0.1 \mu\text{m}^3$  (BL-AC10EGS, Olympus Corporation).

**Surface plasmon resonance (SPR) assay:** SPR experiments were performed using a BIACORE X instrument. The biotinylated ODN 1 (5'-

bitoin-TTTTCGCGGGGCGGGGCGGGGCGGGG-3') were immobilized on a streptavidin-coated sensor chip SA to obtain the desired immobilization levels. Experiments were performed using degassed and filtered HBS (10 mM HEPES, 150 mM NaCl, 3 mM EDTA, and 0.005% surfactant P20) buffer with 100 mM KCl at 25 °C (pH 7.4). Sample solutions at various concentrations were prepared in HBS-EP buffer with 100 mM KCl and injected at a flow rate of 20  $\mu\text{L}/\text{min}$ . Data processing was performed by global fitting of the sensorgrams obtained experimentally to a model of two site binding using the BIAevaluation 4.1 software.

**DNAzyme experiments:** Generally, DNAzyme solution with 1  $\mu\text{M}$  DNA in  $\text{K}^+$  solution (20 mM lithium cacodylate, pH 7.0, 100 mM KCl) was prepared and annealing was performed by heating to 95°C for 5 min and cooling down to room temperature slowly. Then hemin and 2, 2'-azino-di-[3-ethylbenzthiazoline sulfonate (ABTS) were added to the final concentration of 1  $\mu\text{M}$  and 2 mM accordingly. DNAzyme experiments were started by adding 10  $\mu\text{L}$   $\text{H}_2\text{O}_2$  (final concentration 1 mM). Absorption of the solution at 420 nm was collected every 30 seconds from 0 to 60 min a spectrophotometer V-650 (JASCO, Japan) in a time-dependent method.

**Luciferase reporter assay:** The HBM promoter sequences were amplified from human genomic DNA using AccuPrime™ GC-Rich DNA Polymerase (Life Technologies, Japan). The pMCS Cypridina luciferase reporter (Thermo Scientific, Japan) plasmids containing HBM promoter were constructed. HEK293T cells (ATCC, Japan) were seeded in 24-well plates with  $8 \times 10^4$  cells/well. To observe the effect of the G-quadruplex ligands, the plasmid containing the HBM promoter (1  $\mu\text{g}$ ) was transfected into the cells using FuGENE 6 transfection reagent (Promega, Japan). After 2 hr the cells were treated with 1  $\mu\text{M}$  and 10  $\mu\text{M}$  compounds. To understand the role of ATRX, halo-tag ATRX plasmid (Kazusa DNA Research Institute, Japan) and the HBM promoter plasmids were cotransfected by Lipofectamine® 3000 Transfection Reagent (Life Technologies, Japan). The luciferase activity was measured after 48 hr with Pierce Cypridina luciferase Flash Assay Kit (Thermo Scientific, Japan) at 463 nm using a Spectra Max 190 (Molecular Devices, Japan) microplate reader.

## Acknowledgements

This work was partially supported by the Practical Research Project for Rare/Intractable Diseases from Japan Agency for Medical Research and Development, AMED, and JSPS KAKENHI Grant Numbers 24225005. Funding for open access charge: JSPS.

**Keywords:** ATR-X syndrome • ATRX • G-quadruplex • VNTR sequence •  $\alpha$ -globin gene

- [1] Y. Xue, R. Gibbons, Z. Yan, D. Yang, T.L. McDowell, S. Sechi, J. Qin, S. Zhou, D. Higgs, W. Wang, *Proc. Natl. Acad. Sci. U S A.* **2003**, *100*, 10635-10640.
- [2] J. Tang, S. Wu, H. Liu, R. Stratt, O.G. Barak, R. Shiekhathar, D.J. Picketts, X. Yang, *J. Biol. Chem.* **2004**, *279*, 20369-20377.
- [3] P.W. Lewis, S.J. Elsaesser, K.M. Noh, S.C. Stadler, C.D. Allis, *Proc. Natl. Acad. Sci. U S A.* **2010**, *107*, 14075-14080.
- [4] P. Drané, K. Ouarhni, A. Depaux, M. Shuaib, A. Hamiche, A. *Genes. Dev.* **2010**, *24*, 1253-1265.
- [5] A. Corpet, T. Olbrich, M. Gwerder, D. Fink, M. Stucki, *Cell Cycle.* **2014**, *13*, 249-267.

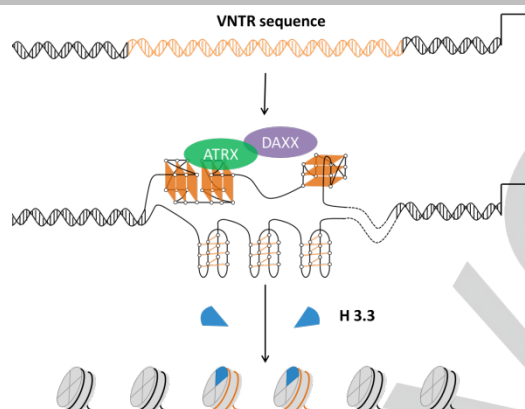


- [6] A. Argentaro, J.C. Yang, L. Chapman, M.S. Kowalczyk, R.J. Gibbons, D.R. Higgs, D. Neuhaus, D. Rhodes, *Proc. Natl. Acad. Sci. U S A*. **2007**, *104*, 11939-11944.
- [7] K. Ratnakumar, E. Bernstein, *Epigenetics*. **2013**, *8*, 3-9.
- [8] S. Iwase, B. Xiang, S. Ghosh, T. Ren, P.W. Lewis, J.C. Cochrane, C.D. Allis, D.J. Picketts, D.J. Patel, H. Li, Y. Shi, *Nat. Struct. Mol. Biol.* **2011**, *18*, 769-776.
- [9] S. Eustermann, J.C. Yang, M.J. Law, R. Amos, L.M. Chapman, C. Jelinska, D. Garrick, D. Clynes, R.J. Gibbons, D. Rhodes, D.R. Higgs, D. Neuhaus, *Nat. Struct. Mol. Biol.* **2011**, *18*, 777-782.
- [10] R.J. Gibbons, D.J. Picketts, L. Villard, D.R. Higgs, *Cell*. **1995**, *80*, 837-845.
- [11] D.J. Picketts, D.R. Higgs, S. Bachoo, D.J. Blake, O.W. Quarrell, R.J. Gibbons, *Hum. Mol. Genet.* **1996**, *5*, 1899-1907.
- [12] Wada, T., Kubota, T., Fukushima, Y., and Saitoh, S. *Am. J. Med. Genet.* **2000**, *94*, 242-248.
- [13] R.J. Gibbons, T. Wada, C.A. Fisher, N. Malik, M.J. Mitson, D.P. Steensma, A. Fryer, D.R. Goudie, I.D. Krantz, J. Traeger-Synodinos, *Hum. Mutat.* **2008**, *29*, 796-802.
- [14] T. Wada, H. Ban, M. Matsufuji, N. Okamoto, K. Enomoto, K. Kurosawa, N. Aida, *AJNR. Am. J. Neuroradiol.* **2013**, *34*, 2034-2038.
- [15] K. Ratnakumar, L.F. Duarte, G. LeRoy, D. Hasson, D. Smeets, C. Vardabasso, C. Bönisch, T. Zeng, B. Xiang, D.Y. Zhang, H. Li, X. Wang, S.B. Hake, L. Schermelleh, B.A. Garcia, E. Bernstein, *Genes. Dev.* **2012**, *26*, 433-438.
- [16] K.D. Kernohan, D. Vernimmen, G.B. Gloor, N.G. Bérubé, *Nucleic Acids Res.* **2014**, *42*, 8356-8368.
- [17] K. Sarma, C. Cifuentes-Rojas, A. Ergun, A. Del Rosario, Y. Jeon, F. White, R. Sadreyev, J.T. Lee, *Cell*. **2014**, *159*, 869-883.
- [18] N. Shioda, H. Beppu, T. Fukuda, E. Li, I. Kitajima, K. Fukunaga, *J. Neurosci.* **2011**, *31*, 346-358.
- [19] R.L. Flynn, K.E. Cox, M. Jeitany, H. Wakimoto, A.R. Bryll, N.J. Ganem, F. Bersani, J.R. Pineda, M.L. Suvà, C.H. Benes, D.A. Haber, F.D. Boussin, L. Zou, *Science*. **2015**, *347*, 273-277.
- [20] D. Clynes, C. Jelinska, B. Xella, H. Ayyub, C. Scott, M. Mitson, S. Taylor, D.R. Higgs, R.J. Gibbons, *Nat. Commun.* **2015**, *6*, 7538.
- [21] S.M. Mirkin, *Nature*. **2007**, *447*, 932-940.
- [22] R. Simone, P. Fratta, S. Neidle, G.N. Parkinson, A.M. Isaacs, *FEBS Lett.* **2015**, *589*, 1653-1668.
- [23] N. Maizels, *EMBO Rep.* **2015**, *16*, 910-922.
- [24] S. Burge, G.N. Parkinson, P. Hazel, A.K. Todd, S. Neidle, *Nucleic Acids Res.* **2006**, *34*, 5402-5415.
- [25] A.R. Haeusler, C.J. Donnelly, G. Periz, E.A. Simko, P.G. Shaw, M.S. Kim, N.J. Maragakis, J.C. Troncoso, A. Pandey, R. Sattler, J.D. Rothstein, J. Wang, *Nature*. **2014**, *507*, 195-200.
- [26] K. Zhang, C.J. Donnelly, A.R. Haeusler, J.C. Grima, J.B. Machamer, P. Steinwald, E.L. Daley, S.J. Miller, K.M. Cunningham, S. Vidensky, S. Gupta, M.A. Thomas, I. Hong, S.L. Chiu, R.L. Haganir, L.W. Ostrow, M.J. Matunis, J. Wang, R. Sattler, T.E. Lloyd, J.D. Rothstein, *Nature*. **2015**, *525*, 56-61.
- [27] A.C. Connor, K.A. Frederick, E.J. Morgan, L.B. McGown, *J. Am. Chem. Soc.* **2006**, *128*, 4986-4991.
- [28] Z. Yu, J.D. Schonhoft, S. Dhakal, R. Bajracharya, R. Hegde, S. Basu, H. Mao, *J. Am. Chem. Soc.* **2009**, *131*, 1876-1882.
- [29] S. Millevoi, H. Moine, S. Vagner, *Wiley Interdiscip. Rev. RNA*. **2012**, *3*, 495-507.
- [30] R.Y. Wu, K.W. Zheng, J.Y. Zhang, Y.H. Hao, Z. Tan, *Angew. Chem. Int. Ed. Engl.* **2015**, *54*, 2447-2451.
- [31] Y. Sannohe, M. Endo, Y. Katsuda, K. Hidaka, H. Sugiyama, *J. Am. Chem. Soc.* **2010**, *132*, 16311-16313.
- [32] G. Biffi, D. Tannahill, J. McCafferty, S. Balasubramanian, *Nat. Chem.* **2014**, *6*, 75-80.
- [33] V. Brázda, L. Hároníková, J.C. Liao, M. Fojta, *Int. J. Mol. Sci.* **2014**, *15*, 17493-17517.
- [34] M.L. Bochman, K. Paeschke, V.A. Zakian, *Nat. Rev. Genet.* **2012**, *13*, 770-780.
- [35] P. Murat, S. Balasubramanian, *Curr. Opin. Genet. Dev.* **2014**, *25*, 22-29.
- [36] T.M. Ou, Y.J. Lu, J.H. Tan, Z.S. Huang, K.Y. Wong, L.Q. Gu, *ChemMedChem*. **2008**, *3*, 690-713.
- [37] S. Balasubramanian, L.H. Hurley, S. Neidle, *Nat. Rev. Drug Discov.* **2011**, *10*, 261-275.
- [38] K. Shinohara, Y. Sannohe, S. Kaieda, K. Tanaka, H. Osuga, Y. Xu, T. Bando, H. Sugiyama, *J. Am. Chem. Soc.* **2010**, *132*, 3778-3782.
- [39] M.J. Law et al, *Cell*. **2010**, *143*, 367-378.
- [40] M.A. Levy, K.D. Kernohan, Y. Jiang, N.G. Bérubé, *Hum. Mol. Genet.* **2015**, *24*, 1824-1835.
- [41] D.R. Higgs, D. Vernimmen, B. Wood, *Adv. Genet.* **2008**, *61*, 143-173.
- [42] S.H. Goh, Y.T. Lee, N.V. Bhanu, M.C. Cam, R. Desper, B.M. Martin, R. Moharram, R.B. Gherman, J.L. Miller, *Blood*. **2005**, *106*, 1466-1472.
- [43] K. Guo, A. Pourpak, K. Beetz-Rogers, V. Gokhale, D. Sun, L.H. Hurley, *J. Am. Chem. Soc.* **2007**, *129*, 10220-10228.
- [44] L. Petraccone, *Top. Curr. Chem.* **2012**, *330*, 23-46.
- [45] H. Martadinata, A.T. Phan, *Biochemistry*. **2013**, *52*, 2176-2183.
- [46] K.J. Neaves, J.L. Huppert, R.M. Henderson, J.M. Edwardson, *Nucleic Acids Res.* **2009**, *37*, 6269-6275.
- [47] I. Mela, R. Kranaster, R.M. Henderson, S. Balasubramanian, J.M. Edwardson, *Biochemistry*. **2012**, *51*, 578-585.
- [48] S. Pisano, M. Varra, E. Micheli, T. Coppola, P. De Santis, L. Mayol, M. Savino, *Biophys. Chem.* **2008**, *136*, 159-163.
- [49] A. Siddiqui-Jain, C.L. Grand, D.J. Bearss, L.H. Hurley, *Proc. Natl. Acad. Sci. U S A*. **2002**, *99*, 11593-11598.
- [50] B. Zamiri, K. Reddy, R.B. Jr. Macgregor, C.E. Pearson, *J. Biol. Chem.* **2014**, *289*, 4653-4659.
- [51] L. Stefan, F. Denat, D. Monchaud, *J. Am. Chem. Soc.* **2011**, *133*, 20405-20415.
- [52] H. Yaku, T. Murashima, D. Miyoshi, N. Sugimoto, *Molecules*. **2012**, *17*, 10586-10613.
- [53] J.C. Grigg, N. Shumayrikh, D. Sen, *PLoS One*. **2014**, *9*, e106449.
- [54] D. Clynes, C. Jelinska, B. Xella, H. Ayyub, S. Taylor, M. Mitson, C.Z. Bachrati, D.R. Higgs, R.J. Gibbons, *PLoS One*. **2014**, *9*, e92915.
- [55] D. Clynes, R.J. Gibbons, *Curr. Opin. Genet. Dev.* **2013**, *23*, 289-294.
- [56] D. Clynes, D.R. Higgs, R.J. Gibbons, *Trends Biochem. Sci.* **2013**, *38*, 461-466.
- [57] C. Sissi, B. Gatto, M. Palumbo, *Biochimie*. **2011**, *93*, 1219-1230.
- [58] K. Paeschke, M.L. Bochman, P.D. Garcia, P. Cejka, K.L. Friedman, S.C. Kowalczykowski, V.A. Zakian, *Nature*. **2013**, *497*, 458-462.

## Entry for the Table of Contents

## FULL PAPER

A model of ATRX and G-quadruplex formation was proposed and it might explain the vital disease-causing role of different size of VNTR sequence and ATRX mutations found in  $\alpha$ -thalassemia of ATR-X syndrome.



Yue Li,<sup>[a]</sup> Junetha Syed,<sup>[a]</sup> Yuki Suzuki,<sup>[a]</sup> Sefan Asamitsu,<sup>[a]</sup> Norifumi Shioda,<sup>[c]</sup> Takahito Wada,<sup>[a]</sup> and Hiroshi Sugiyama,<sup>\*,[a, b]</sup>

Page No. – Page No.

Effect of ATRX and G-quadruplex formation by the VNTR sequence on  $\alpha$ -globin gene expression



# SUPPORTING INFORMATION

## Effect of ATRX and G-quadruplex formation by the VNTR sequence on $\alpha$ -globin gene expression

Yue Li,<sup>[a]</sup> Junetha Syed,<sup>[a]</sup> Yuki Suzuki,<sup>[a]</sup> Sefan Asamitsu,<sup>[a]</sup> Norifumi Shioda,<sup>[c]</sup> Takahito Wada,<sup>[d]</sup> and Hiroshi Sugiyama,<sup>\*,[a, b]</sup>

[a] Y. Li, S. Asamitsu, Dr. J. Syed, Dr. Y. Suzuki, Prof. H. Sugiyama

*Department of Chemistry, Graduate School of Science, Kyoto University, Kyoto 606-8502, Japan*

*E-mail: hs@kuchem.kyoto-u.ac.jp*

[b] Prof. H. Sugiyama

*Institute for Integrated Cell-Material Sciences (WPI-iCeMS), Kyoto University, Kyoto 606-8501, Japan*

[c] Dr. N. Shioda

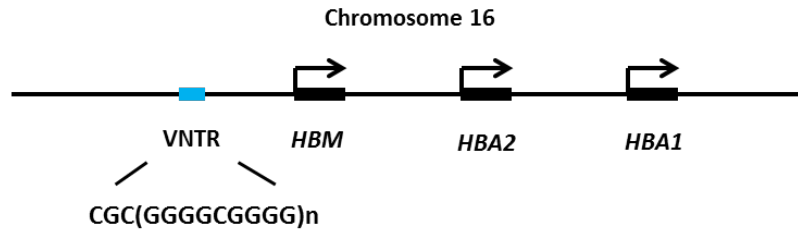
*Department of Pharmacology, Graduate School of Pharmaceutical Sciences, Tohoku University, Sendai 980-8578, Japan*

[d] Prof. T. Wada

*Department of Medical Ethics and Medical Genetics, Graduate School of Medicine, Kyoto University, Kyoto 606-8501, Japan*

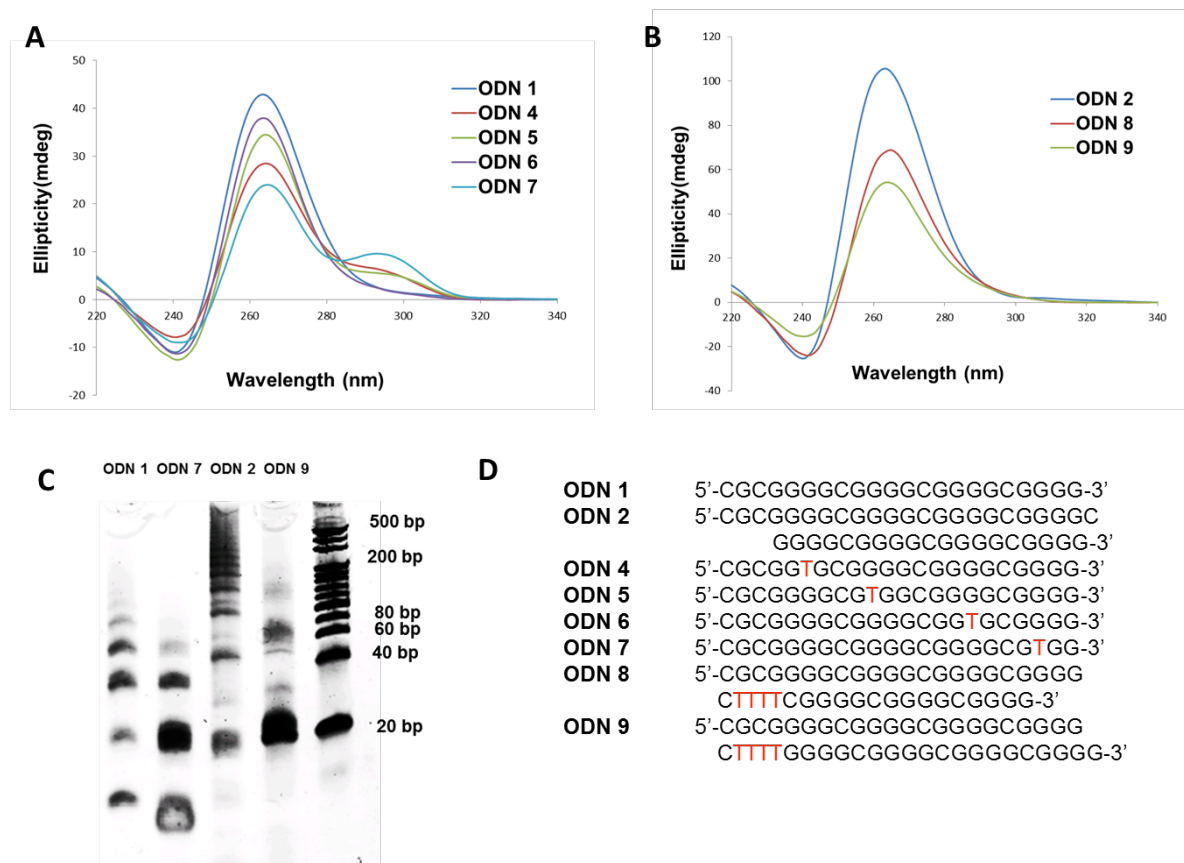
#### Oligonucleotides and primers used in this study

**ODN 1** 5'-CGCGGGGCGGGGCGGGGCGGGG-3' (22mer)  
**ODN 2** 5'-CGCGGGGCGGGGCGGGGCGGGGCGGGGCGGGGCGGGGCGGGGCGGGG-3' (42mer)  
**ODN 3** 5'-CCCCGCCCCGCCCCGCCCCGCG-3' (22mer)  
**ODN 4** 5'-CGCGGTGCGGGGCGGGGCGGGG-3' (22mer)  
**ODN 5** 5'-CGCGGGGCGTGGCGGGGCGGGG-3' (22mer)  
**ODN 6** 5'-CGCGGGGCGGGGCGGTGCGGGG-3' (22mer)  
**ODN 7** 5'-CGCGGGGCGGGGCGGGGCGTGG-3' (22mer)  
**ODN 8** 5'-CGCGGGGCGGGGCGGGGCGGGGCTTTTCGGGGCGGGGCGGGG-3' (42mer)  
**ODN 9** 5'-CGCGGGGCGGGGCGGGGCGGGGCGGGGCTTTTGGGGCGGGGCGGGGCGGGG-3' (46mer)  
**Biotin-ODN 1** 5'-bitoin-TTTTCGCGGGGCGGGGCGGGGCGGGG-3'  
**c-myc** 5'-GGGGAGGGTGGGGAGGGTGGGGAAGG-3'  
**telomere** 5'-AGGGTTAGGGTTAGGGTTAGGG-3'  
**Forward primer for 1.4 kb** 5'-TACATCCTGCGCGTGGACCCGGT-3'  
**Forward primer for 0.9 kb** 5'-CTGTCCCACTGCCTGCTGGTCA-3'  
**Reverse primer** 5'-AGGTCTGGGACCTGATTGCGG-3'



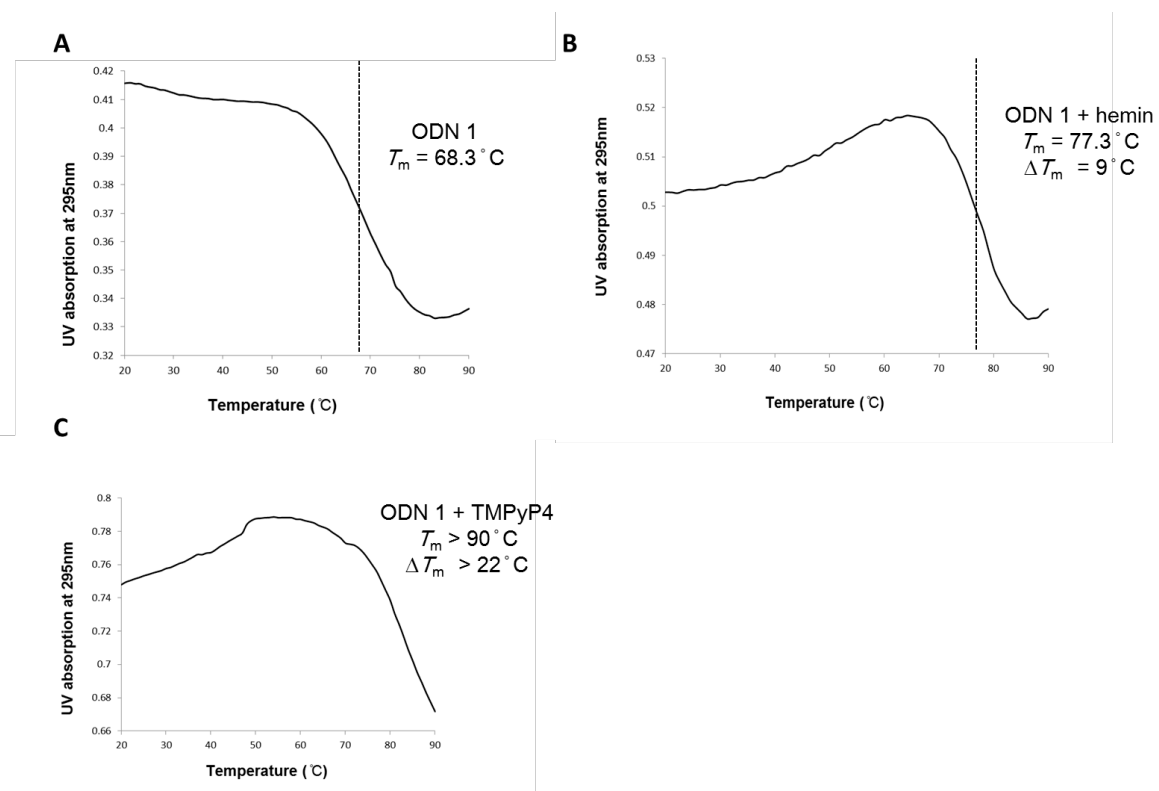
CAAGGTGCGCGGGGCGCGGTGCGGGCGGGGCGGGGCGGGGCCGCGGGGCGGGCGGGGCCGC  
 GGGGCGGGGTCGCGGGGCGGGGCGGGGTGGGGTCGCGGGGCGGGGCGGGGTCGCGGGGCG  
 GGGCGGGGCGGGGCGGGGCGGGGCGGGGCGGCCGGGGCCGCGGGGCGGGGCGGGGCGGG  
 GAGGGGCTGGGCGGGGCGGGGCGCGGGGCGGGGCGGGCCGGGGCCGGGGCGGGGTCGCGGG  
 GCGGGGTCGCGGGGCGGGGCGCGGGGCGGGGCGGGGCGGGGTGGGGTCGCGGGGCGGGG  
 CCGGGCTA

**Figure S1.** VNTR sequence was located in the promoter region of *HBM* gene in the  $\alpha$ -globin gene cluster. It contains hundreds of CGC(GGGGCGGG)n repeats and the size of VNTR sequence is variable among individuals. Here is an example of whole VNTR sequence from the data of *HBM* gene. It mainly consists of CGC(GGGGCGGG)n repeats with several mutations.



**Figure S2.** Mutation studies of G-quadruplex formation by VNTR sequence. CD spectra (A and B) and gel analysis (C) of ODN 1, ODN 2 and mutated ODNs (D).





**Figure S3.** UV melting temperature of ODN 1 (A), ODN 1 with hemin (B) and ODN 1 with TMPyP4 (C). Oligonucleotides (5  $\mu\text{M}$ ) in  $\text{K}^+$  solution (20 mM lithium cacodylate, pH 7.0, 5 mM KCl) and 1 eq compounds were added. The melting temperature of ODN 1 with TMPyP4 was higher than 90  $^\circ\text{C}$  in 5 mM KCl, indicating the high stabilization ability of TMPyP4 towards G-quadruplex formed by ODN 1.

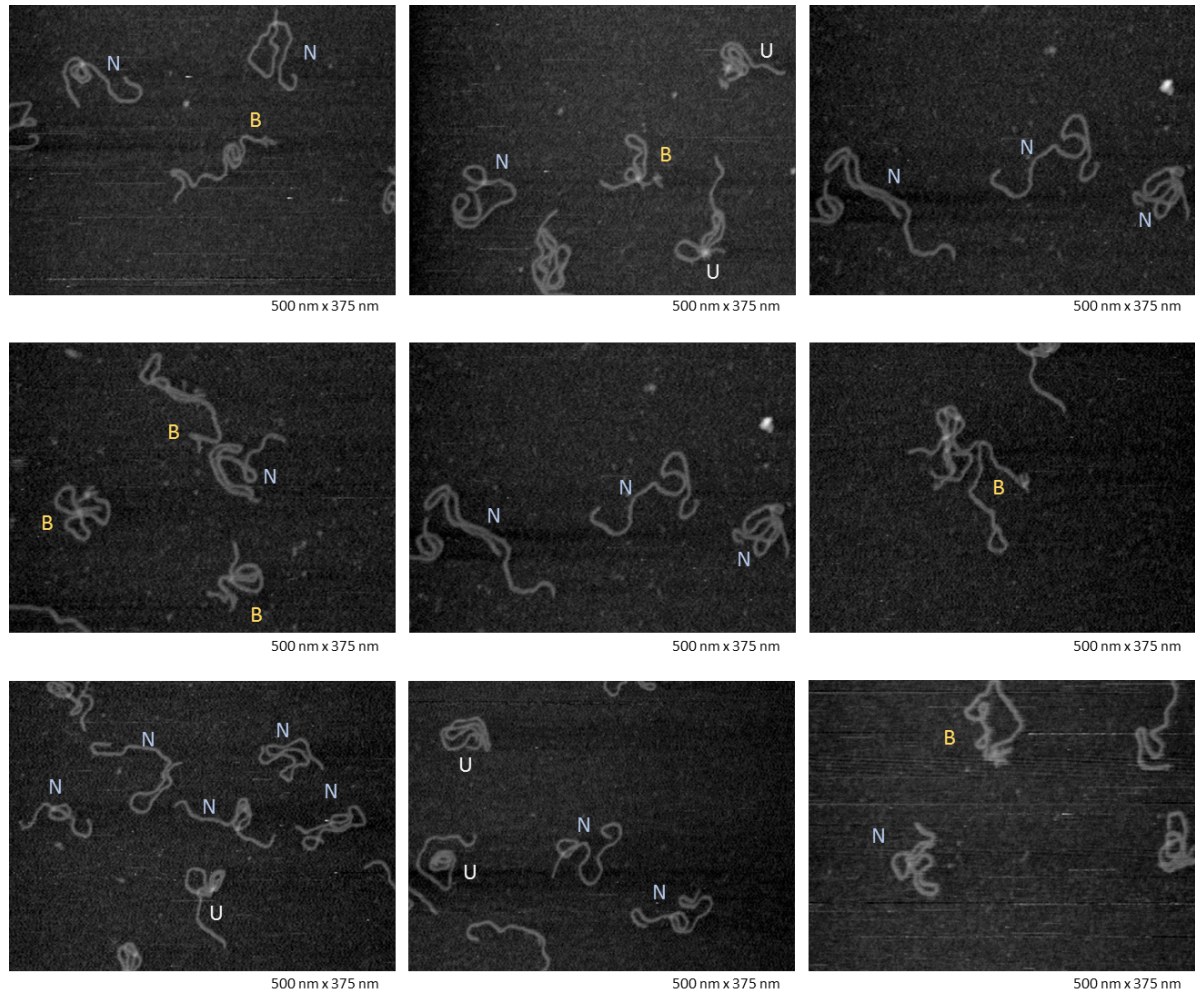
*HBM* promoter

1.4 kb

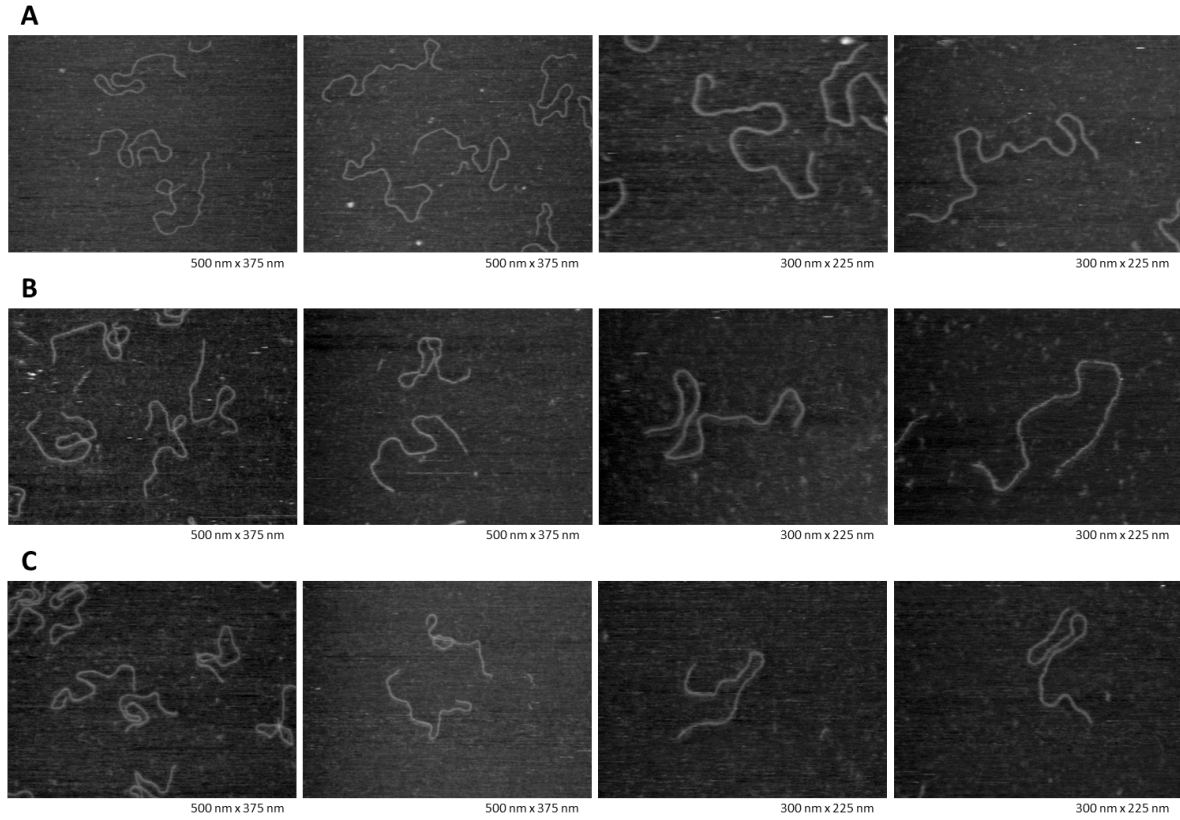
TACATCCTGCGCGTGGACCCGGTCAACTTCAAGGTCGCGGGGCGCGGTCGGGGCGGGGCGGGGCGGGGCGGGGCGGGGCGGGG  
CGGGGCGCGGGGCGGGGTCGCGGGGCGGGGCGGGGTGGGGTCGCGGGGCGGGGCGGGGTCGCGGGGCGGGGCGGGGCGGGG  
CGGGGCGGGGCGGGGCGGGCCGGGCGCGGCGGGGCGGGGCGGGGAGGGGCTTGGGCGGGGCGGGGCGGGGCGGGGCGGGG  
CGGGCGGGGCGGGGCGGGGTCGCGGGGCGGGTCGCGGGGCGGGGCGGGGCGGGGCGGGGCGGGGTCGGGTCGCGGGG  
CGGGGCGGGGCTAGGCCCGCCCGCACTGAGCGCCCCCGCCAGCTCCTGTCCCATGCTGCTGCTCACCTGGCTGGTCCACCTGG  
CCGCGCGCTTCCCGCGCGACTTACGCGCGAGGCCACGCCGCTGGGCCAAGTTCCTATCGTTCGTATCCTCTGTCTCTGA  
CCGAGAAGTACCGCTGAGCGCCGCTC CGGGACCCCGAGGACAGGCTGCGGCCCTCCCTGCCCTTACCCCTCCACAGT  
TCTTGCCCTGACTCCAATAAATGGATGAGGACGGAGCGATCTGGGCTCTGTGTTCTCAGTATTGGAGGGAAGGAGGGGAGA  
AGCTAGCTGATGGGTCCGGGGCTTCGAGGAATCGGTCTGCCCATGTCTGCTCGCGGCTGGGTTCACCTGGGGGGCG  
CCTTGGGGAGGTTCTAGCCCTGAGCACCGGAGCTGCGGCCGGGTGGAGCGGAGCAGTCCCGGGCGGGCCGCGGCGTCT  
CCTGGGGTCTTGAGTCGGACGGGCGTTTGTGCGTCTCCCGCTTCCCATATCGCACAAGATTGTCACTTCACTAAGCGT  
ATTGGAAGCGTGTGCGGGCTCAGGGAACCTTTCCACAAAGCTGACGTCCGAATCCCGGACTCTGGCAGCTACGGGGTCT  
CCTGAGGCGGGTCCCTCCCGACTCCTAAGAGGTAGGGGGTTTCTGCGCGGTGTCTCTCTCGGTTCTCCCATGTGC  
TCCCTCCTGGCAGAGCAGTAACCTTACCCGAGGGGAGTAAACAGATGCCCTAAAGTCTGCAGTAAAGTGCCACGCGCA  
ACGGCGTGGGTCAATGCCAGAAACCTGGGATCCCGGAGGTCGAGGCTCCACACAGACGGGAACCGGGCTGGTTACGTT  
CCCCGGCGAGGCGAGGGCTCCCGCGTTCCCGCCGCGCTCGGGCCGATAAGGACGGGCGGGGTGCCCGGAGGCTCTATAA  
GGAGGCCAGGCGGCGGGCGGGCCCCAGAGCAGTCAGGCGGCGCATGCTCAGCGCCAGGAGCGGCCCAATCGCG  
CAGGTCCTGGGACCTGATTCCGG

0.9 kb

**Figure S4.** 1.4 kb promoter sequence and 0.9 kb promoter sequence used in the AFM experiments and luciferase reporter assay.



**Figure S5.** AFM images of the 1.4 kb *HBM* promoter sequence were obtained at pH 5.5 in 100 mM KCl solution. Three types of structures were identified: B (branched structure), N (non-branched structure) and U (unidentified structure). The percentage of branched structures is 22% (14/63).

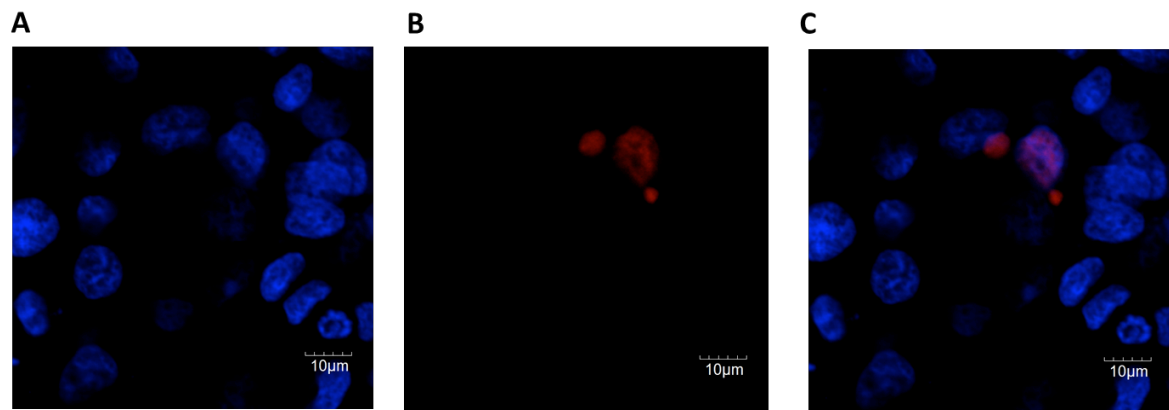


**Figure S6.** A) AFM images of 1.4 kb sequence in pH 5.5 without  $K^+$ . B) AFM images of 1.4 kb sequence in pH 7.6 with 100 mM  $K^+$ . C) AFM images of 0.9 kb sequence without VNTR G-quadruplex-forming sequence. In all three cases, no branched structures were visualized, suggesting no G-quadruplex or i-motif formation.



Compounds	$K_{D1}$	$K_{D2}$	$K_D$	$\Delta T_m$
Hemin	$8.5 \times 10^{-5}$ M	$6.2 \times 10^{-6}$ M	$2.3 \times 10^{-5}$ M	9 °C
TMPyP4	$2.3 \times 10^{-7}$ M	$1.6 \times 10^{-8}$ M	$6.1 \times 10^{-7}$ M	>22 °C

**Table S1.** Binding affinity of hemin and TMPyP4 towards G-quadruplex formed by ODN 1. Two-site binding model (black line) was used to compare with TMPyP4.  $K_D$  ( $2.3 \times 10^{-5}$  M) was calculated by  $K_{D1}$  ( $8.5 \times 10^{-5}$  M) and  $K_{D2}$  ( $6.2 \times 10^{-6}$  M) based on two-site binding model. Compared with TMPyP4 ( $K_D$   $6.1 \times 10^{-7}$  M), hemin has a larger  $K_D$  ( $2.3 \times 10^{-5}$  M), which is consistent with the melting experiment in Figure S3.



**Figure S7.** Expression of ATRX in the HEK293T cells. A) Staining of HEK293T cells with DAPI. B) Staining of HEK293T cells with HaloTag® TMR Ligand. C) Merge of DAPI and HaloTag® TMR Ligand staining. Halo-tag ATRX plasmid (Kazusa DNA Research Institute, Japan) was transfected into HEK293T cells by Lipofectamine® 3000 Transfection Reagent (Life Technologies). After 2 days, cells were stained with DAPI and HaloTag® TMR Ligand (G8251, Promega) following the product instructions. Then cell images were obtained by FV1200 Laser Scanning Microscopes (Olympus, Japan).