Analysis of the Function of the Loops in G-quadruplex in the Promoter of the Transcription Factor BmPOUM2 in the Silkworm, *Bombyx Mori*

Yanfei Chen
South China Normal University - Shipai Campus: South China Normal University
https://orcid.org/0000-0003-0452-5261

Kangkang Niu
South China Normal University - Shipai Campus: South China Normal University

Qisheng Song
University of Missouri

Qili Feng (qlfeng@scnu.edu.cn)
Guangdong Provincial Key Laboratory of Insect Developmental Biology and Applied Technology, Institute of Insect Science and Technology, School of Life Sciences, South China Normal University, Guangzhou 510631, China; https://orcid.org/0000-0003-2105-6831

Research

Keywords: G-quadruplex, formation, protein binding, bombyx mori

Posted Date: July 28th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-739638/v1

Abstract

Bachguound: A G4 structure in the promoter of the transcription factor *BmPOUM2* was identified in *Bombyx mori.* This G4 structure binds with the transcription factor BmLARK and is involved in the transcription of *BmPOUM2*. However, the structure and functional domains of the *BmPOUM2* G4 remain to be clarified.

Results: In this study, the relationship between the loop structure and the function of the *BmPOUM2* G4 was examined using mutation, CD analysis, EMSA, MST and luciferase assay. The results revealed that the loop mutations could not suppress the G4 structure formation, but loop C does influence the LARK binding affinity and the promoter activity.

Conclusions: the results of this study are suggesting that the structure and function of *BmPOUM2* G4 are mainly determined by the G-quartets, which maintain the framework of the structure and were targeted by BmLARK. Nevertheless, the loops, which are relatively variable in G4, may entitle the variability of the structure, especially the loop C, which is the one-nucleotide-loop.

Background

In addition to the double helix, DNA molecules often contain secondary structures, such as G-quadruplexes (G4s), in which four guanines form a G-quartet and two or more tandem G-quartets linked by the loops form a G4 structure [1]. G4s can form within or between DNA molecules [1]. A DNA G4 sequence would spontaneously fold to be a G4 structure in the course of replication or transcription and is involved in the regulation of these processes [2]. The G4 folding and unfolding are reversible depending on the cellular status, such as pH and Na^+/K^+ concentration [3, 4].

The formation of G4s mainly depends on the bonding between the four guanines that form the G-quartets, which are linked by the loops of nucleotides. In other words, a functional G4 structure consists of G-quartets, which are the mainframe of the structure, and the loops, which were proved to provide the variability of the structure and the targets for drugs [5]. Loops play a key role in the overall folding and stability of the G-quadruplexes. The length and sequence of these loops can either stabilize or destabilize a G-quadruplex, because of the interactions between the G-quartets and the loops [6, 7].

With the mutation analysis, it was found that the size of the loops was important in the most favored folding pattern determination of the *VEGF* G4 [8]. The loop length seemed to be more important than the loop sequence. In RNA G4, shorter loops have greater influences on the structural features than the longer loops [9]. An antiparallel or mixed-type hybrid DNA G4 comprises at least a loop with a total length of more than five nucleotides to maintain the structure [10]. In a study on free energy using explicit solvent molecular dynamics, it was found that the short loops were generally less stable than the longer ones and might bring distortion or instability to the stem structure of G4 [11]. The G4 in the *hTERT* core promoter showed unexpected stability because of the formation of a hairpin loop with 26 bases [12]. Loops were

reported to determine the formation and function of the *BCL2* G4 [13]. In the study of the binding between the Arg-Gly-Gly repeat (RGG) of Ewing's sarcoma protein (EWS) and G4s, Takahama et al. found that the RGG of EWS binds preferentially to G4 with longer loops, regardless of the topology of the G4 structure and the DNA sequence [14].

In our previous study, we found a G4 structure located at -120 to -94 nts upstream of the promoter of the transcription factor *BmPOUM2* [15], which is involved in metamorphosis in *Bombyx mori* [16]. This G4 structure was bound by a conserved transcription factor BmLARK [4]. The two RNA recognition motif (RRM) domains in BmLARK were found to be necessary for the binding with the *BmPOUM2* G4 [17]. The BmPOUM2 G4 consists of four G-quartets and three loops that link the G-quartets, and the loops were named as loop A, loop B and loop C respectively, from the 5'- to 3'-end. Loop A consists of 5 nucleotides, Loop B consists of 5 nucleotides and Loop C has 1 nucleotide. When the guanines in the G4 structure were mutated, BmLARK failed to bind with the G4 structure and the *BmPOUM2* promoter activity was suppressed [15]. However, little is known about the effects of the loop of the G4 structure on the formation of G4 structure, the binding between BmLARK and the G4 structure and function in the regulation of the BmPOUM2 promoter activity.

In this paper, the structural importance and function of the loop elements of the *BmPOUM2* G4 were studied using different approaches. The results indicated that in the case of *BmPOUM2* G4, the loop sequences are of minor importance in the formation, the binding with the acting factor and the function of the G4 structure, except loop C, which is a single nucleotide loop.

Results

The spacial structure of the BmPOUM2 G4 and designs of the loop mutations

Base on our previous work, the possible schematic diagram of the three-dimensional configuration of *BmPOUM2* G4 was shown in Figure1A. Three loops from the 5'-end to 3'-end and the four G-quartets were predicted. To study their function, the nucleotides of the loops were mutated according to the principle of purine to purine, pyrimidine to pyrimidine individually or in combination, to minimize the change to the structure (Figure 1B). All the mutations were predicted by the website: http://bioinformatics.ramapo.edu/QGRS/index.php[18] to assess the formation of the G4 structure and evaluated by G-score as shown in Table 1. From the G-scores of the mutant types, the G4 structure was formed as the wild type.

Table 1. Nucleotides of the wild-type and loop mutants of the *BmPOUM2* G4

Name[1]	QGRS[2]	G-Score[3]
G mut	**GATG**CGCGA**GTGA**GCCGA**GTAG**C**AGTG**	0
WT	**GGGG**CGCGA**GGGG**GCCGA**GGGG**C**GGGG**	59
Loop A mut	GGGG**TATA**AGGGGGCCGAGGGGCGGGG	59
Loop B mut	GGGGCGCGAGGGG**ATTA**AGGGGCGGGG	59
Loop C mut	GGGGCGCGAGGGGGCCGAGGGG**T**GGGG	59
Loop ABC mut	GGGG**TATA**AGGGG**ATTA**AGGGG**T**GGGG	59
Loop AB mut	GGGG**TATA**AGGGG**ATTA**AGGGGCGGGG	59
Loop AC mut	GGGG**TATA**AGGGGGCCGAGGGG**T**GGGG	59
Loop BC mut	GGGGCGCGAGGGG**ATTA**AGGGG**T**GGGG	59

[1] "G mut" represents the sequence in that the guanines in G-quartets were mutated; "WT" represents the wild-type G4 sequence contained in the promoter region of the gene *BmPOUM2* G4; and "A/B/C mut" represents the corresponding loop(s) mutated. [2] Nucleotide sequences of the G4 motif in the wild type and mutated loops. Nucleotides in G-quartet regions are indicated by an underline. The mutated nucleotides are in red bold. [3] G-Score represents the value of possible folding of the corresponding G4 structure predicted by the QGRS Mapper.

The effect of loop mutation on the G4 structure formation

A typical circular dichroism (CD) spectrum of a G4 structure includes an enhanced feature absorption peak at 265 nm and an enhanced feature trough at 240 nm when 100 mM K^+ is present, as compared to the K^+ absence samples [19, 20]. If the nucleotides in the G-quartets were mutated, although the absorption peak and trough were still present, the wavelength of the peaks and troughs were shifted from 265 nm and 240 nm to 270 nm and 250 nm, respectively (Figure 2A), indicating the G4 sequence could not form the G4 structure [21]. In this study, in all the cases of mutations, when 100 mM K^+ was present, absorption peak and trough were enhanced and the CD spectrum patterns were similar to the wild types (Figure 2B-2H), although the values were different, suggesting that the G4 structure was formed in these oligonucleotides. In our previous work [15], the enhanced absorption peak for the wild-type G4 structure was about 5 mdge at 265 nm, and the enhanced absorption trough was about -2 mdge at 240 nm (Figure 2A). However, the enhanced absorption peaks of the mutants of Loop A, B and C (Figure 2B-D), as well as the double mutants of loop A and C (Figure 2F) and the double mutants of loop B and C (Figure 2G), were about or even over than 6 mdge, while the enhanced absorption peaks of the double mutants of loop A and B (Figure 2E) and triple mutants of loop A, B and C (Figure 2H) remained about 5 mdge as the wild type. Whereas the enhanced absorption troughs of the loop mutated sequences were all close to -4 mdge, which were lower than the wild type. The differential changes in the CD values in these loop mutants

might reflect the different stability of the G4 structure with the mutated nucleotides in the loop motifs. However, the G4 formation of the loop mutations appeared to have not been affected.

The effect of G4 loop mutations on binding with BmLARK.

The above experiments indicated that the nucleotide mutants in the loops of the G4 structure did not seriously impact the formation of the G4 structure. To determine if the mutations in the loops influence the binding of BmLARK with the G4 structure, Electrophoretic Mobility Shift Assays (EMSAs) were performed using the G4s with different loop mutations and BmLARK protein, which contained the two RRM domains sufficiently required for the binding with the *BmPOUM2* G4. When no BmLARK protein was present, all the mutants, either Loop A, B and C mutated individually, or double mutants in loop A and B, loop A and C, and loop B and C, or triple mutants in loop A, B and C, formed the G4 structure (Figure 3A). The loop C mutant showed an extra band of the G4 structure (Figure 3A, Lane 5). When BmLARK was present, the BmLARK-G4 binding bands were found in all the mutants (Figure 3B). The bindings between BmLARK and the loop C mutant (Figure 3B, Lane 5) was seemly to be even stronger than the wild type (Figure 3B, Lane 2), and the other binding bands were weaker but clear. When 100 times of the unlabeled competitive probes were added, the BmLARK-G4 binding bands were much weakened or disappeared (Figure 3C).

In all the cases, the BmLARK protein bound the *BmPOUM2* G4 with the loop A, B or C mutated in a similar binding pattern to the wild type G4 (Figure 3B), but the binding affinity appeared to be weakened, as compared to the wild type, suggesting that the loop mutation could not completely block the binding of the protein to the G4 structure, but have an impact on the binding affinity. It also implies that the target of the binding of BmLARK with the G4 structure probably is on the G-quartet(s), but not on the loops.

To further analyze the binding affinity of BmLARK and the *BmPOUM2* G4 with the loops mutated, microscale thermophoresis (MST) tests, which are commonly used for testing the binding affinity between biological molecules [22], were also carried out(Figure 4). The results showed that all the loop mutants, either single or double or triple loop mutants, showed the binding characteristic curves similar to the wild type *BmPOUM2* G4 (Figure 4A), but were distinct from the G-quartets mutated sequence (Figure 4B), which had a muchhigher *Kd* value with BmLARK than the wild-type and othermutants. Almost all the loop mutants have close *Kd* values, of which the *Kd* values of the loop C mutant (Figure 4E), and the loop ABC mutant (Figure 4F) were lower than the wild type. The *Kd* values of the loop AC mutant (Figure 4H) and the loop BC mutant (Figure 4I) were slightly higher or lower than the wild type, and the *Kd* values of the loop A mutant (Figure 4C), of the loop B mutant (Figure 4D) and of the loop AB mutant (Figure 4G) are somewhat higher than the wild type, but the difference values are much smaller than that between the wild type and the G mutant type. The *Kd* values were indicating that the binding affinity between BmLARK and the *BmPOUM2* G4 of G-quartet mutant was much weaker than the wild type and the loop mutated types. And what is more, comparing the binding affinity with BmLARK, loop C mutant and loop ABC mutant were a little stronger than the wild type, the mutant type of loop AC and loop BC were more close to the wild type, and the loop A, loop B and loop AB were a little weaker than the wild type. These results

indicate that the loop mutations did have some impact on the BmLARK-G4 binding affinity, but the effect was not as vital as the G-quartet mutation (Figure 4) and the differences between the loop mutants and the wild type were apparently existing.

The effect of G4 loop mutation on the BmPOUM2 transcription.

To determine whether the loop mutations of the *BmPOUM2* G4 impact the transcription of the gene, luciferase activity analysis was conducted. The promoter region of -169 to +74 nts, containing the G4 sequence and the core promoter, was cloned into the pGL3-basic expression vector, which expresses the marker gene luciferase. The loop nucleotides were mutated in the same way as described above (Table 1).

The results showed that when the G-quartets were mutated, the transcription activity of the BmPOUM2 gene was reduced as compared to the wild type (Figure 5). When Loop A or B was mutated, the transcription activity was not changed. When Loop C was mutated, the transcription activity was enhanced. When both loop A and C or both loop B and C were mutated, the transcription activities were different from the wild type with a little lower (loop AC mutant) or a little higher (loop BC mutant) gene expression. When loop A and B or loop A, B and C were mutated simultaneously, the transcription activity was almost the same as the wild type, with the mutants of loop AB and loop ABC non-significantly lower. In general, mutations in the loop nucleotides did not significantly change the transcription activity of the *BmPOUM2* promoter, but the mutation in the G-quartets suppressed the transcription activity of the gene.

Discussion

As an important transcription factor, BmPOUM2 regulates the expression of *BmWCP4*, which is involved in silkworm metamorphosis [16]. In our previous study, The transcription of *BmPOUM2* is regulated by its G4 structure in the promoter region [15]. The G4 structure failed to form and the promoter activity was reduced obviously when the G-quartet nucleotides were mutated [15]. The transcription factor BmLARK bound the G4 structure of the *BmPOUM2* and regulated the gene expression [4, 17]. However, it is not clear whether the changes in the loop structure of the *BmPOUM2* G4 can affect the G4 formation, the binding of BmLARK and G4, and the transcription activity of the target gene. In this paper, we studied the effects of the loop mutations on the G4 formation and transcription activity of *BmPOUM2* by using CD analysis, EMSA, MST and promoter activity analysis, keeping the length and size of the mutated loops were kept identical to the original ones to avoid extra interference.

Firstly, we found that although the sequences with the loop mutations can still form the G4 structure (Figure 2) and the formed G4 structure can still bind to BmLARK (Figure 3), the binding affinity (Figure 3 and 4) and the transcriptional activities (Figure 5) were affected to a different extend. This suggests that the loops in the entire G4 structure may have some space effect on the binding of BmLARK and the G4 structure. In the study on structure stability and thrombin binding of the G4 formed by thrombin-binding DNA aptamer (TBA), it was demonstrated that thrombin stabilizes the G4 via the interaction with

nucleotides in the loops rather than via direct stabilization of G-quartets [23]. Loop structure and function of G4s were studied in the proto-oncogenes in human beings by replacing guanine with 8-oxo-G mutation to alter the length, topology or the site of the loops, and it was found that the G4 loops provide regulatory function in the enzymatic activation of the gene *PARP-1* in DNA damage repairing [24].

Secondly, we found that in the mutant (cytosine substituted by thymine) of the loop C, which has only one nucleotide cytosine, the polymorphism of the G4 structure and the binding with BmLARK (Figure 3B) appeared to be enhanced (Figure 3A and 3B, Lane 5, Figure 4E). And the combinational mutations with loop C, ie loop AC mutant, loop BC mutant and loop ABC mutant, impacted the gene transcription, positively or negatively, more effectively. In the G4 structure, it is common that the one-nucleotide-loop could enhance the polymorphism of the structure and possess more functional influence, directly or indirectly. A single G-to-A replacement in the G4 loops affected the entire telomere structure [25]. It was reported that there are some common interesting sequence characteristics for G4 and G4 that can be inhibited by simply substituting thymine with adenosine [26]. In the study of the structures and stability of G4 formed in *Myc1234*, the different base conformations of the single nucleotide loops would cause changes in the thermostability of *c-MYC* G4 [27]. The C-to-A mutation in the loop C of the *BmPOUM2* G4 probably was in favor of the formation of the G4 and its binding with BmLARK.

The alternation or substitution with the tested nucleotides in the loops of *BmPOUM2* G4, especially one-nucleotide-loop, loop C, can affect the binding affinity with BmLARK and the transcriptional activity of the promoter. As long as the G-quartets were holding the special structure, no matter how the loops were mutated, either single loop mutants, double loops mutants or triple loops mutants, the G4 can still be formed, as indicated by the typical *cis*-G4 absorption spectrum (Figure 2), and The G4 binding protein BmLARK still bound the G4 structure with these loop mutants as well. Nevertheless, with merely the loop, the formation of the G4 structure(Figure 2A), the binding with the binding protein(Figure 3A), and the transcriptional activity (Figure 4) were significantly suppressed[15], indicating that the loops were not sufficient for the structure and function of the G4, but were necessary.

Conclusions

In this study, we found that, when the loop(s) was (were) mutated, the sequence of *BmPOUM2* G4 can still form the stereochemical structure of G4 and the stereochemical structure can still bind to the binding protein BmLARK. In the case of substitution of loop A or loop B nucleotides did not significantly change the G4 structure, the binding with the binding protein BmLARK, and the transcription activity of the *BmPOUM2* promoter. However, the binding affinity and the function of the promoter were interfered in the case of loop C mutation. So that, we hypothesize that the binding site of BmLARK is mainly on the G-quartet planes, which maintained the spatial structure of G4, and some loops of the G4, particularly the one nucleotide loops, may be responsible for the variability of G4 (Figure 6).

Materials And Methods

The probes used in the assays

For CD analysis, the probes of the wild-type, G-quartets mutants and the loop mutants of the *BmPOUM2* G4 were synthesized by Tsingke (Shanghai, China).

For EMSA experiments, the oligonucleotide probes were labeled by biotin (the competitive probes were unlabeled) at the 5'-ends and synthesized by Tsingke (Shanghai, China). For MST experiments, the oligonucleotide probes were labeled with Cy5 at the 5'-ends and synthesized by Sangon Biotech Co Ltd. (Shanghai, China). The probes for EMSA and MST assays consisted of seven nucleotides ahead and six nucleotides behind the G4 region to ensure the folding of the secondary structure, for example, the probe for EMSA of the wild type was 5'-BIO- CCAGTGCGGGGCGCGAGGGGGCCGAGGGGCGGGGCAACAA-3' (the G4 region was underlined).

In the luciferase assay for the influence of the loop mutations on the promoter activity, the *BmPOUM2* promoter was shortened to a 243 basepairs fragment (from -169 to +74), which included the G4 motif and the start site for the gene transcription, and synthesized by Tsingke (Shanghai, China).

Cell line and protein expression and purification

A *B. mori* cell line *BmE* originally developed from embryos was provided by the State Key Laboratory of Silkworm Genome Biology (Southwest University, China) and cultured in Grace medium (Invitrogen, USA) supplemented with 10% fetal bovine serum (Gibco, USA).

The open reading frame of the truncated 166 amino acid BmLARK protein (LARK 166), containing the G4-binding domain, was cloned into the pET-28a vector between the *Nde* I and *Xho* I sites with a His-tag at the N-terminus of the protein. The expression of the recombinant protein was induced by 1 mmol/L isopropyl beta-D-thiogalactopyranoside (IPTG). The protein was then purified with Ni-NTA His·Bind Resin (Millipore, USA) and eluted using elution buffer (0.5 mol/L NaCl, 20 mmol/L Tris-HCl, 1 mol/L imidazole, pH7.9). The concentration of the protein was determined with BCA Protein Assay Reagent Kit (Thermo Scientific, USA).

CD analysis

Five μM DNA oligonucleotides containing the G4 sequence and 2 extra nucleotides at both ends of the G4 sequence to ensure the structure formation were incubated in 50 mM Tris buffer (pH 7.5) with 0 mM or 100 mM KCl at 95°C for 10 min and then cooled to room temperature slowly for more than 4 h to facilitate the G4 structure formation. CD analysis was carried out in a J-815 CD spectrometer (Jasco International, USA) at the wavelength from 220 to 350 nm with 1 nm step width and 1 s response time. Three averaged scans of the same sample were taken to standardize the CD spectra.

EMSA

The oligonucleotides of the wild-type and all the mutants were heated at 95 °C for 10 min in Tris buffer at pH 7.5 with or without 100 mM KCl and then allowed to gradually cool to room temperature to facilitate the G4 structure formation. Then the G4-BmLARK binding reactions were conducted using a Light Shift Chemiluminescent EMSA Kit (Thermo Scientific, MA, USA). The binding mixture contained 1µl each of 50% glycerol, 1% NP-40, 1 M KCl, 100 mM $MgCl_2$, 200 mM EDTA, and 2 µl of 10×binding buffer (10 mM Tris, 50 mM KCl, 1 mM DTT, pH 7.5), 20 fmol DNA biotinylated probe, and 1 µg LARK 166 in a total volume of 20 µl. For the competitive reactions, an additional 4 µl of 50 µmol unlabeled same oligonucleotide probes were added. All the mixtures were incubated at room temperature for 20 min to complete the binding reaction.

After the binding reaction, the binding was examined in 4% polyacrylamide gel followed by electrotransfering the samples to a positively charged Nylon membrane (Amersham Biosciences, Boston, USA). The membranes were photographed on a ChemiDoc™ Touch Imaging System (BioRad, USA).

MST

After denatured in 50 mM Tris buffer at pH 7.5 with 100 mM KCl at 95°C for 10 min, the 5'-end Cy5 labeled G4 oligonucleotides were cooled to room temperature for more than 4 h to allow the G4 structure formation. The purified BmLARK protein (LARK 166) was diluted into different concentrations from 2.93 to 3,000 nM and mixed with 25 nM labeled oligonucleotides in the binding buffer [20 mM Tris (pH 7.5), 100 mM NaCl and 0.05% Tween-20] and incubated for 20 min at room temperature.

The incubated mixtures were then loaded into the corresponding standard capillaries (Nanotemper, Germany) and analyzed on the Monolith NT.115 (NanoTemper Technologies GmbH, Munich, Germany) at 24°C with LED 60% and 40% MST power. Three technical repeats were performed for each of the samples. The data analysis was done by using the software of NT Control and NT Analysis.

Luciferase assay for promoter activity analysis

To examine the effect of mutation of the loop nucleotides on the *BmPOUM2* promoter activity, the recombinant pGL3-derived vectors containing the 243bp wild type or mutated promoter (Table 1) were used to determine the luciferase activity. The internal control plasmid pRL-SV40 (Promega, USA), which contains renilla luciferase activity, was used as the reference plasmid. Ten µl co-transfection reagent containing 200 ng recombinant plasmid, 10 ng pRL-SV40, 0.6 µl Fugene HD transfection reagent (Promega, USA) in the Opti-MEM Reduced Serum Medium (Life Technologies, USA) was added to *BmE* cells in the 48-well plate and cultured for overnight (12 h). Each sample of the pGL3-derived vectors was repeated in three cell culture wells. The luciferase activity measurement was conducted using Dual Luciferase Reporter Gene Assay Kit (Yeasen, Shanghai China). In short, the culture medium was removed and the cells were washed twice with PBS. 150µl cell lysis buffer was added to each well to lyse the cells for 5 min. The cell samples were collected and centrifuged at 12,000rpm for 1 min. Ten µl of the supernatant was used to measure the luciferase activity in the firefly luciferase and renilla luciferase working solution in a luminometer (IBA7300, Veritas, Turner Biosystems).

All assays were repeated at least three times. The luciferase activity was represented as mean ± standard error (SE). Statistical significance of the luciferase activity was analyzed using Student's t-test in GraphPad Prism7.

Abbreviations

G4 G-quadruplex

CD Circular Dichroism

EMSA Electrophoretic Mobility Shift Assay

MST microscale thermophoresis

Declarations

Acknowledgements

This work was supported by the grants of Chinese National Natural Science

Foundation(Grant no.: 31720103916, 31930102).

Authors' contributions

YC conducted most of the experiments and the data analyses and drafting the manuscript. KN participated in the project design, helped drafting and revising the manuscript. QS participated drafting and revising the manuscript. QF conceived, designed and oversaw the research, provided financial support, drafted and finalized the manuscript. All authors read and approved the final manuscript.

Funding

This research was funded by the grants of Chinese National Natural Science Foundation (Grant no.: 31720103916, 31930102).

Availability of data and materials

Data sharing not applicable to this article as no datasets were generated or analyzed during the current study.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

References

1. Burge, S.; Parkinson, G. N.; Hazel, P.; Todd, A. K.; Neidle, S., Quadruplex DNA: sequence, topology and structure. *Nucleic Acids Res* **2006,** *34* (19), 5402-15.

2. Maizels, N.; Gray, L. T., The G4 genome. *PLoS Genet* **2013,** *9* (4), e1003468.

3. Wang, C.; Jia, G.; Li, Y.; Zhang, S.; Li, C., Na+/K+ switch of enantioselectivity in G-quadruplex DNA-based catalysis. *Chem Commun (Camb)* **2013,** *49* (95), 11161-3.

4. Niu, K.; Xiang, L.; Jin, Y.; Peng, Y.; Wu, F.; Tang, W.; Zhang, X.; Deng, H.; Xiang, H.; Li, S.; Wang, J.; Song, Q.; Feng, Q., Identification of LARK as a novel and conserved G-quadruplex binding protein in invertebrates and vertebrates. *Nucleic Acids Research* **2019**.

5. Song, J. H.; Kang, H. J.; Luevano, L. A.; Gokhale, V.; Wu, K.; Pandey, R.; Sherry Chow, H. H.; Hurley, L. H.; Kraft, A. S., Small-Molecule-Targeting Hairpin Loop of hTERT Promoter G-Quadruplex Induces Cancer Cell Death. *Cell Chem Biol* **2019,** *26* (8), 1110-1121 e4.

6. Huppert, J. L., Balasubramanian, S., Loop-length-dependent folding of G-quadruplexes. *Nucleic Acids Research* **2004**, 2908-2916.

7. Prislan, I.; Khutsishvili, I.; Marky, L. A., Interaction of minor groove ligands with G-quadruplexes: thermodynamic contributions of the number of quartets, T-U substitutions, and conformation. *Biochimie* **2011,** *93* (8), 1341-50.

8. Guo, K.; Gokhale, V.; Hurley, L. H.; Sun, D., Intramolecularly folded G-quadruplex and i-motif structures in the proximal promoter of the vascular endothelial growth factor gene. *Nucleic Acids Res* **2008,** *36* (14), 4598-608.

9. Zhang, A. Y.; Bugaut, A.; Balasubramanian, S., A sequence-independent analysis of the loop length dependence of intramolecular RNA G-quadruplex stability and topology. *Biochemistry* **2011,** *50* (33), 7251-8.

10. Bugaut, A.; Balasubramanian, S., A sequence-independent study of the influence of short loop lengths on the stability and topology of intramolecular DNA G-quadruplexes. *Biochemistry* **2008,** *47* (2), 689-97.

11. Cang, X.; Sponer, J.; Cheatham, T. E., 3rd, Insight into G-DNA structural polymorphism and folding from sequence and loop connectivity through free energy analysis. *J Am Chem Soc* **2011,** *133* (36), 14270-9.

12. Palumbo, S. L.; Ebbinghaus, S. W.; Hurley, L. H., Formation of a unique end-to-end stacked pair of G-quadruplexes in the hTERT core promoter with implications for inhibition of telomerase by G-quadruplex-interactive ligands. *J Am Chem Soc* **2009,** *131* (31), 10878-91.

13. Dai, J.; Chen, D.; Jones, R. A.; Hurley, L. H.; Yang, D., NMR solution structure of the major G-quadruplex structure formed in the human BCL2 promoter region. *Nucleic Acids Res* **2006,** *34* (18), 5133-44.

14. Takahama, K.; Sugimoto, C.; Arai, S.; Kurokawa, R.; Oyoshi, T., Loop lengths of G-quadruplex structures affect the G-quadruplex DNA binding selectivity of the RGG motif in Ewing's sarcoma. *Biochemistry* **2011,** *50* (23), 5369-78.

15. Niu, K.; Zhang, X.; Deng, H.; Wu, F.; Ren, Y.; Xiang, H.; Zheng, S.; Liu, L.; Huang, L.; Zeng, B.; Li, S.; Xia, Q.; Song, Q.; Palli, S. R.; Feng, Q., BmILF and i-motif structure are involved in transcriptional regulation of BmPOUM2 in Bombyx mori. *Nucleic Acids Res* **2018,** *46* (4), 1710-1723.

16. Deng, H.; Zhang, J.; Li, Y.; Zheng, S.; Liu, L.; Huang, L.; Xu, W. H.; Palli, S. R.; Feng, Q., Homeodomain POU and Abd-A proteins regulate the transcription of pupal genes during metamorphosis of the silkworm, Bombyx mori. *Proceedings of the National Academy of Sciences* **2012,** *109* (31), 12598-12603.

17. Peng, Y.; Niu, K.; Yu, G.; Zheng, M.; Wei, Q.; Song, Q.; Feng, Q., Identification of binding domains and key amino acids involved in the interaction between BmLARK and G4 structure in the BmPOUM2 promoter in Bombyx mori. *Insect Sci* **2020**.

18. Kikin, O.; D'Antonio, L.; Bagga, P. S., QGRS Mapper: a web-based server for predicting G-quadruplexes in nucleotide sequences. *Nucleic Acids Res* **2006,** *34* (Web Server issue), W676-82.

19. David, A. P.; Margarit, E.; Domizi, P.; Banchio, C.; Armas, P.; Calcaterra, N. B., G-quadruplexes as novel cis-elements controlling transcription during embryonic development. *Nucleic Acids Research* **2016,** *44* (9), 4163-4173.

20. Banerjee, K.; Wang, M.; Cai, E.; Fujiwara, N.; Baker, H.; Cave, J. W., Regulation of tyrosine hydroxylase transcription by hnRNP K and DNA secondary structure. *Nature Communications* **2014,** *5* (1).

21. David, A. P.; Margarit, E.; Domizi, P.; Banchio, C.; Armas, P.; Calcaterra, N. B., G-quadruplexes as novel cis-elements controlling transcription during embryonic development. *Nucleic Acids Res* **2016,** *44* (9), 4163-73.

22. Wienken, C. J.; Baaske, P.; Rothbauer, U.; Braun, D.; Duhr, S., Protein-binding assays in biological liquids using microscale thermophoresis. *Nat Commun* **2010,** *1*, 100.

23. Nagatoishi, S.; Isono, N.; Tsumoto, K.; Sugimoto, N., Loop residues of thrombin-binding DNA aptamer impact G-quadruplex stability and thrombin binding. *Biochimie* **2011,** *93* (8), 1231-8.

24. Edwards, A. D.; Marecki, J. C.; Byrd, A. K.; Gao, J.; Raney, K. D., G-Quadruplex loops regulate PARP-1 enzymatic activation. *Nucleic Acids Res* **2021,** *49* (1), 416-431.

25. Miyoshi, D.; Karimata, H.; Sugimoto, N., Drastic effect of a single base difference between human and tetrahymena telomere sequences on their structures under molecular crowding conditions. *Angew Chem Int Ed Engl* **2005,** *44* (24), 3740-4.

26. Ric, A.; Ecochard, V.; Iacovoni, J. S.; Boutonnet, A.; Ginot, F.; Ong-Meang, V.; Poinsot, V.; Paquereau, L.; Couderc, F., G-quadruplex aptamer selection using capillary electrophoresis-LED-induced fluorescence and Illumina sequencing. *Anal Bioanal Chem* **2018,** *410* (7), 1991-2000.

27. Mathad, R. I.; Hatzakis, E.; Dai, J.; Yang, D., c-MYC promoter G-quadruplex formed at the 5'-end of NHE III1 element: insights into biological relevance and parallel-stranded G-quadruplex stability. *Nucleic Acids Res* **2011,** *39* (20), 9023-33.

Figures

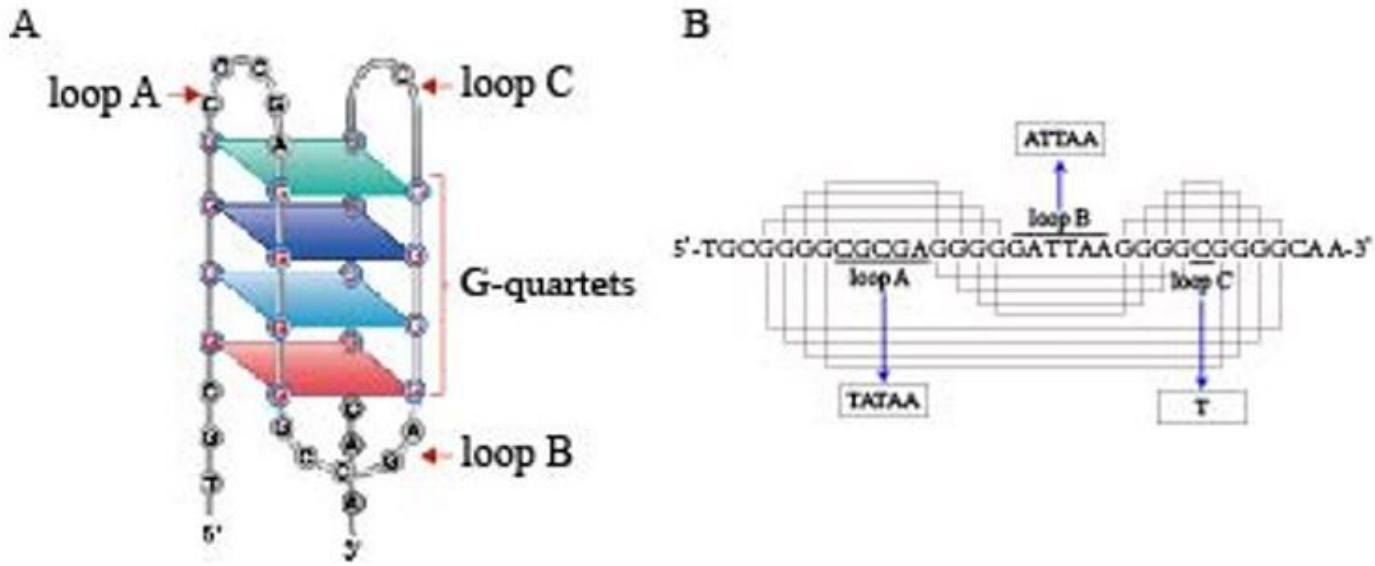

Figure 1

(A) Schematic diagram of three-dimensional configuration of the BmPOUM2 G4. The three loop motifs were named loop A, loop B and loop C, respectively, as shown with red arrow and the four G-quartets were

shown as well. (B) The nucleotide sequences of the BmPOUM2 G4 loop mutants, in which the indicated nucleotides within a loop were mutated simultaneously.

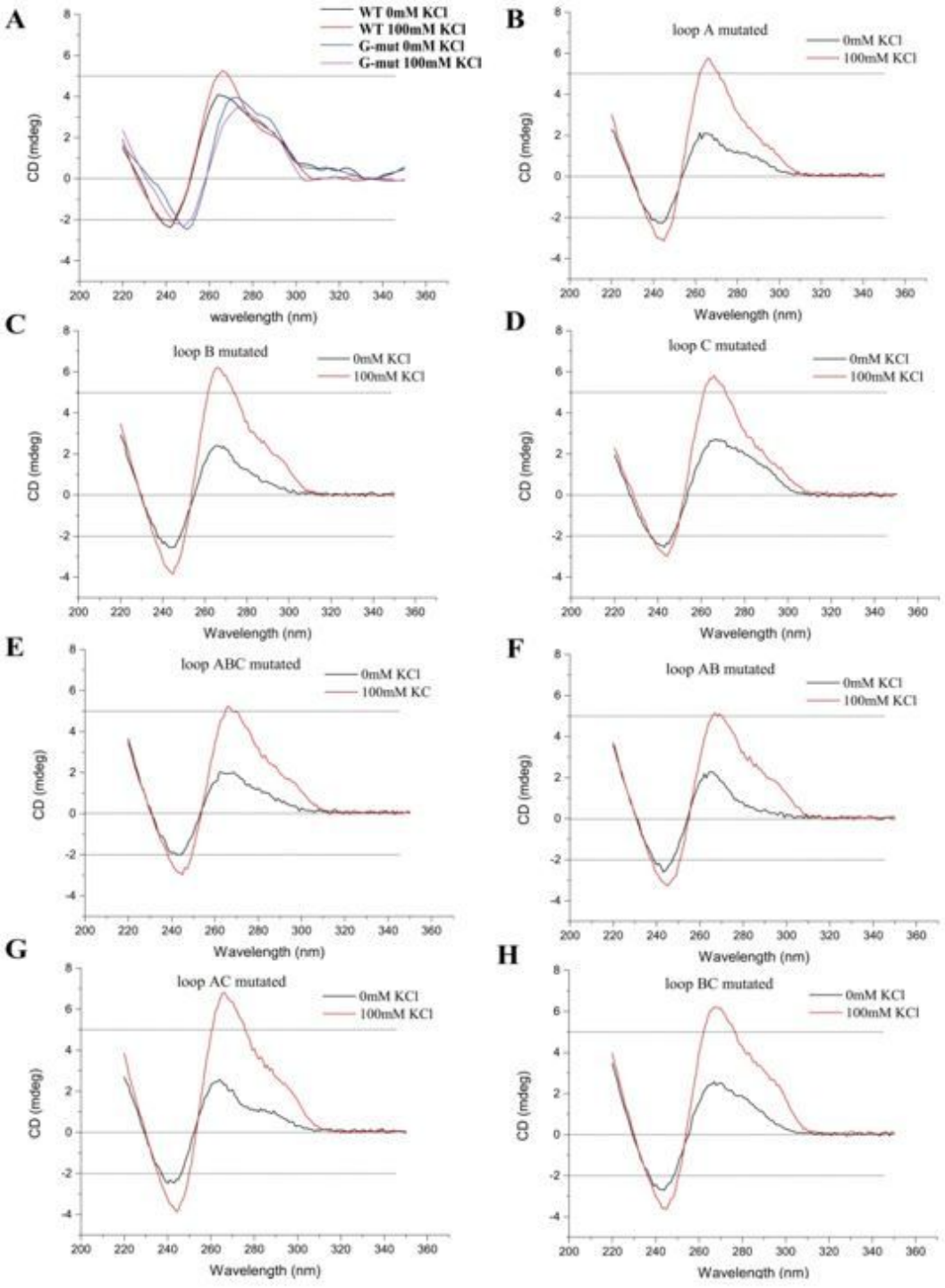

Figure 2

The CD spectra of the BmPOUM2 G4 structures with mutated loop sequences in the presence or absence of 100 mM K+. (A) The CD spectra of the wild type and the G-quartets mutated BmPOUM2 G4. (B) The CD spectra of the loop A-mutated BmPOUM2 G4. (C) The CD spectra of the loop B-mutated BmPOUM2 G4.

(D) CD spectra of loop C-mutated BmPOUM2 G4. (E) The CD spectra of the loop A and loop B mutated BmPOUM2 G4. (F) The CD spectra of the loop A and loop C mutated BmPOUM2 G4. (G) The CD spectra of the loop B and loop C mutated BmPOUM2 G4. (H) The CD spectra of the loop A, B and C mutated BmPOUM2 G4.

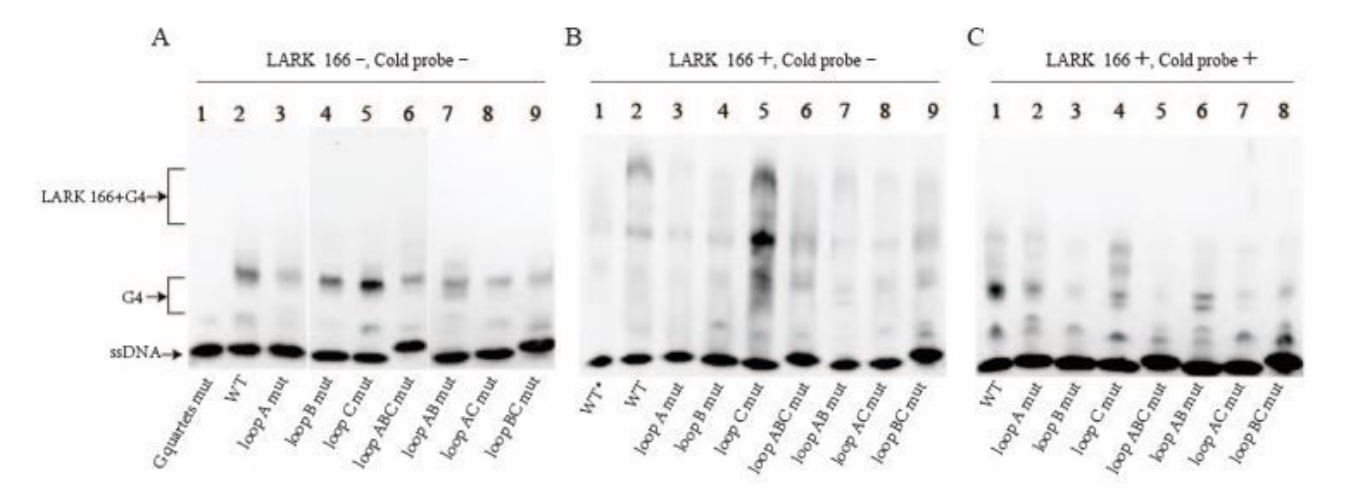

Figure 3

EMSA of the binding between BmLARK (LARK 166) and the BmPOUM2 G4 with different mutated loops. G quartets mut, G quartets mutated BmPOUM2 G4; WT, wild type of BmPOUM2 G4 ; WT*, wild type that was not annealed to fold G4; loop A mut, loop A mutated BmPOUM2 G4; loop B mut, loop B mutated BmPOUM2 G4; loop C mut, loop C mutated BmPOUM2 G4; loop AB mut, loop A and B mutated BmPOUM2 G4; loop AC mut, loop A and C mutated BmPOUM2 G4; loop BC mut, loop B and C mutated BmPOUM2 G4; loop ABC mut, loop A and B and C mutated BmPOUM2 G4. (A) EMSA of the labeled BmPOUM2 G4 probes, which were annealed at room temperature to form the G4 structure, but without the presence of BmLARK protein and cold probe. (B) EMSA of the binding of BmLARK and the labeled BmPOUM2 G4 probes, which were annealed at room temperature to form the G4 structure, in the presence of BmLARK. (C) EMSA of the binding of the BmLARK and the labeled BmPOUM2 G4 probes, which were annealed at room temperature to form the G4 structure, in the presence of BmLARK and the unlabeled cold probes. The band positions of the G4 sequence-containing DNA probes (ssDNA), the G4 structure-containing DNA probes (G4), and the G4-BmLARK bindings are marked by arrows. All the reactions were conducted in the presence of 100 mM K+.

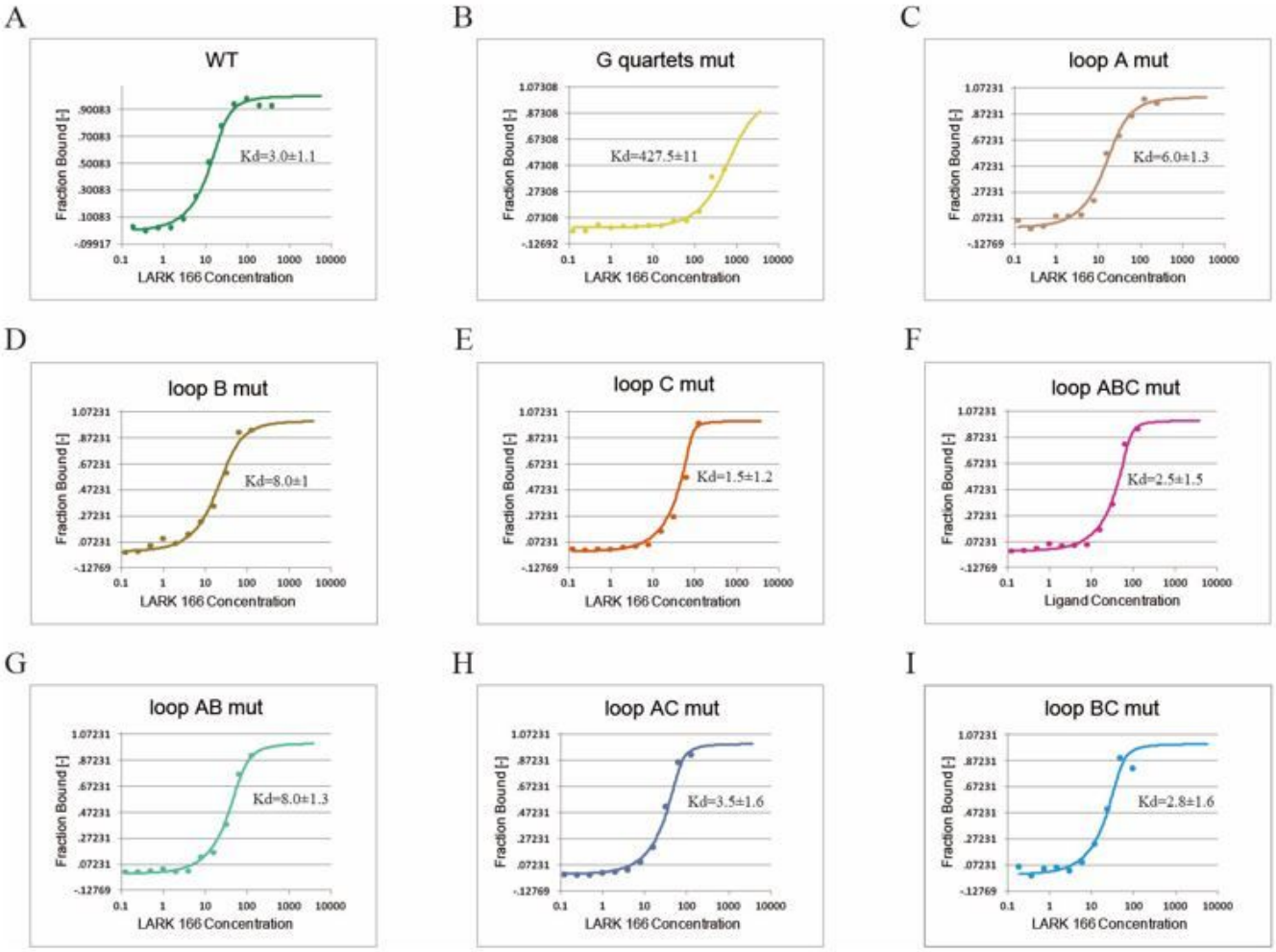

Figure 4

MST measurements of the binding affinity of BmLARK and the BmPOUM2 G4 with different mutants. (A) The binding curves and the Kd values of wild type of BmPOUM2 G4 ; (B) The binding curves and the Kd values of G quartets mutated BmPOUM2 G4; (C) The binding curves and the Kd values of loop A mutated BmPOUM2 G4; (D) The binding curves and the Kd values of loop B mutated BmPOUM2 G4; (E) The binding curves and the Kd values of loop C mutated BmPOUM2 G4; (F) The binding curves and the Kd values of loop A and B mutated BmPOUM2 G4; (G) The binding curves and the Kd values of loop A and C mutated BmPOUM2 G4; (H) The binding curves and the Kd values of loop B and C mutated BmPOUM2 G4; (I) The binding curves and the Kd values of loop A and B and C mutated BmPOUM2 G4.

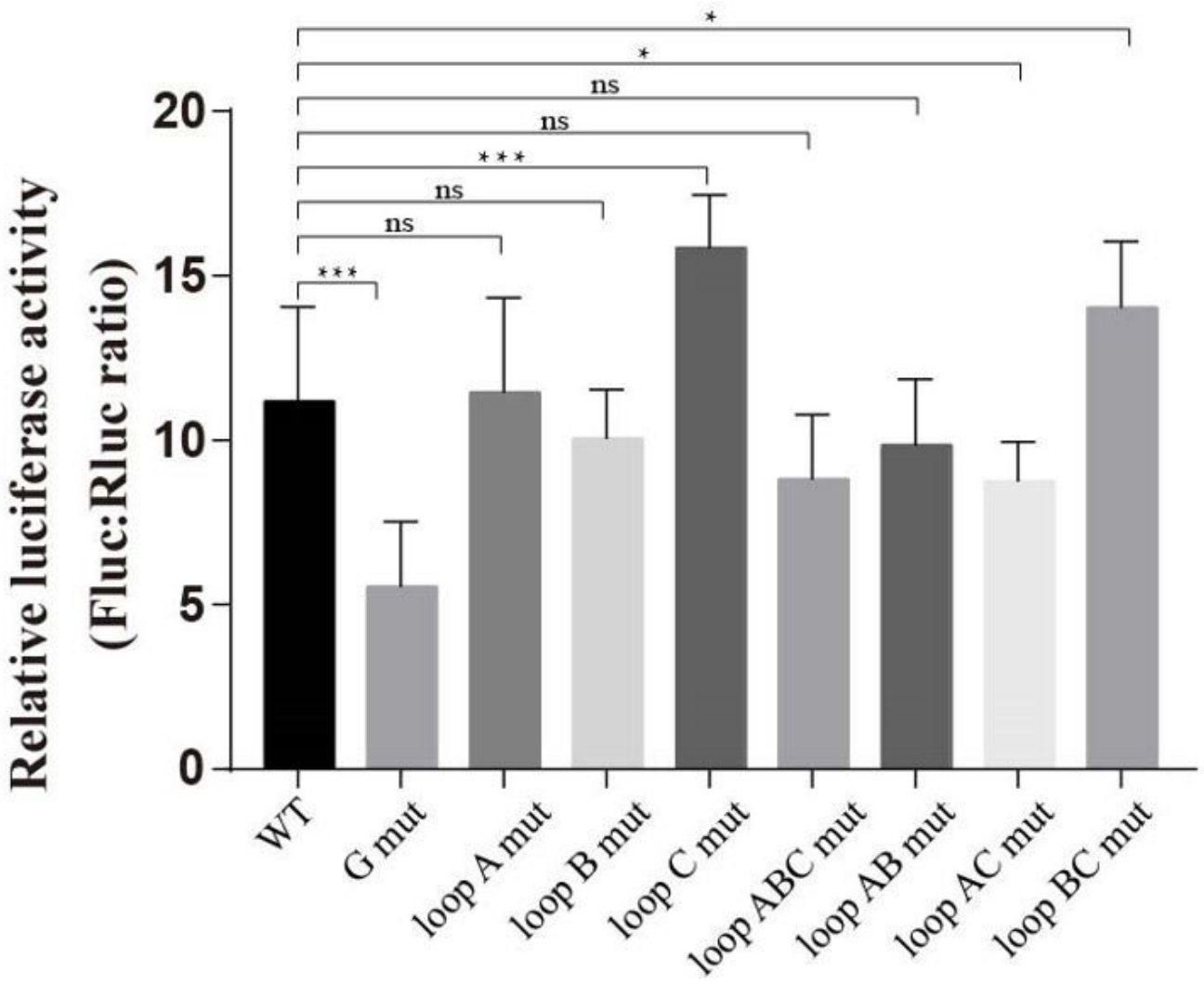

Figure 5

The effect of the loop mutations on the promoter activity of BmPOUM2. WT, wild type; G mut, G-quartets mutated BmPOUM2 G4; loop A mut, loop A mutated BmPOUM2 G4; loop B mut, loop B mutated BmPOUM2 G4; loop C mut, loop C mutated BmPOUM2 G4; loop AB mut, loop A and B mutated BmPOUM2 G4; loop AC mut, loop A and C mutated BmPOUM2 G4; loop BC mut, loop B and C mutated BmPOUM2 G4; loop ABC mut, loop A and B and C mutated BmPOUM2 G4. *p<0.05; **p<0.001; ***P<0.0001. ns, non-significant.

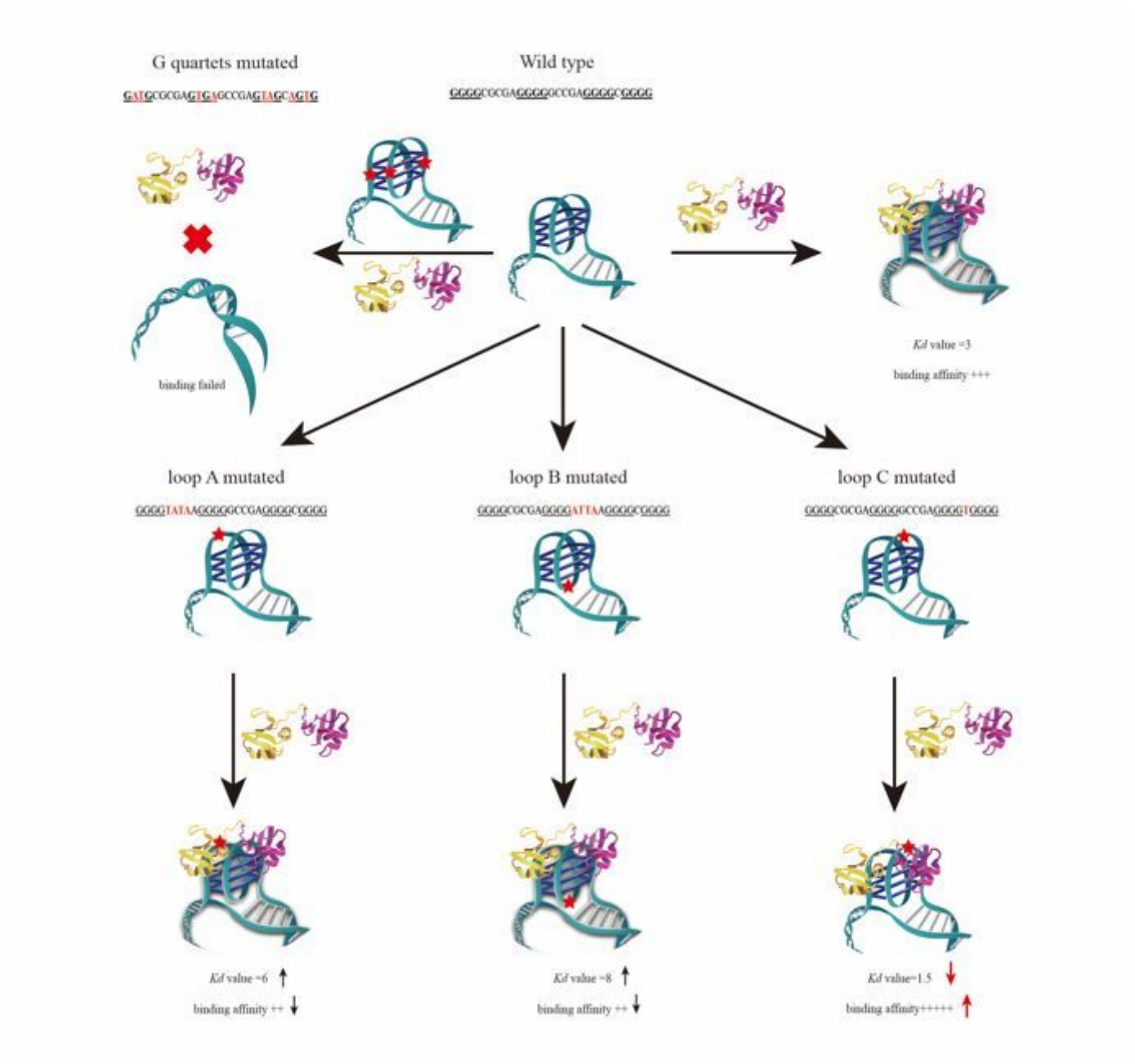

Figure 6

The model of formation and function of the original and mutated BmPOUM2 G4s. In the case of BmPOUM2 G4, when the nucleotides in the G-quartets were mutated, it would be failed to fold to be G4, so that, the binding of BmLARK, which binds to the G4 structure, would be failed. When the nucleotides in the loops were mutated, independently or in combination, the sequence can still form to be G4, and bind to BmLARK as well. Nevertheless, the binding affinities between BmLARK and loop mutants were varied, so that, the activities of the promoter were affected.