2-28-2018

BmILF and I-motif Structure Are Involved in Transcriptional Regulation of *BmPOUM2* in *Bombyx mori*

Kangkang Niu
South China Normal University, China

Xiaojuan Zhang
South China Normal University, China

Huimin Deng
South China Normal University, China

Feng Wu
South China Normal University, China

Yandong Ren
Chinese Academy of Sciences, China

See next page for additional authors

Authors
Kangkang Niu, Xiaojuan Zhang, Huimin Deng, Feng Wu, Yandong Ren, Hui Xiang, Sichun Zheng, Lin Liu, Lihua Huang, Baojuan Zeng, Sheng Li, Qingyou Xia, Qisheng Song, Subba R. Palli, and Qili Feng

BmILF and I-motif Structure Are Involved in Transcriptional Regulation of *BmPOUM2* in *Bombyx mori*

Notes/Citation Information
Published in *Nucleic Acids Research*, v. 46, issue 4, gkx1207, p. 1710-1723.

Digital Object Identifier (DOI)
https://doi.org/10.1093/nar/gkx1207

1710–1723 *Nucleic Acids Research, 2018, Vol. 46, No. 4* *Published online 29 November 2017*
doi: 10.1093/nar/gkx1207

BmILF and i-motif structure are involved in transcriptional regulation of *BmPOUM2* in *Bombyx mori*

Kangkang Niu[1], Xiaojuan Zhang[1], Huimin Deng[1], Feng Wu[1], Yandong Ren[2], Hui Xiang[1], Sichun Zheng[1], Lin Liu[1], Lihua Huang[1], Baojuan Zeng[1], Sheng Li[1], Qingyou Xia[3], Qisheng Song[4], Subba Reddy Palli[5] and Qili Feng[1,*]

[1]Guangzhou Key Laboratory of Insect Development Regulation and Application Research, Institute of Insect Science and Technology, School of Life Sciences, South China Normal University, Guangzhou 510631, China, [2]State Key Laboratory of Genetic Resources and Evolution, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming 650223, China, [3]State Key Laboratory of Silkworm Genome Biology, Southwest University, Chongqing 400716, China, [4]Division of Plant Sciences, University of Missouri, Columbia, MO 65211, USA and [5]Department of Entomology, University of Kentucky, Lexington, KY 40546-0091, USA

Received September 07, 2017; Revised November 12, 2017; Editorial Decision November 20, 2017; Accepted November 23, 2017

ABSTRACT

Guanine-rich and cytosine-rich DNA can form four-stranded DNA secondary structures called G-quadruplex (G4) and i-motif, respectively. These structures widely exist in genomes and play important roles in transcription, replication, translation and protection of telomeres. In this study, G4 and i-motif structures were identified in the promoter of the transcription factor gene *BmPOUM2*, which regulates the expression of the wing disc cuticle protein gene (*BmWCP4*) during metamorphosis. Disruption of the i-motif structure by base mutation, antisense oligonucleotides (ASOs) or inhibitory ligands resulted in significant decrease in the activity of the *BmPOUM2* promoter. A novel i-motif binding protein (BmILF) was identified by pull-down experiment. BmILF specifically bound to the i-motif and activated the transcription of *BmPOUM2*. The promoter activity of *BmPOUM2* was enhanced when BmILF was overexpressed and decreased when BmILF was knocked-down by RNA interference. This study for the first time demonstrated that BmILF and the i-motif structure participated in the regulation of gene transcription in insect metamorphosis and provides new insights into the molecular mechanism of the secondary structures in epigenetic regulation of gene transcription.

INTRODUCTION

Transcriptional factors of the POU domain family, including Pit, Oct, and Unc, have been identified from many species and reported to play critical roles in many developmental processes (1–3). In mammals, POU transcription factors regulate the development of the neuroendocrine systems (4). In insects, POU transcription factor plays a critical role in the neuronal lineage and wiring (5), cell proliferation and differentiation in the developing wing imaginal disc (6), regulation of sericin-1 gene (7), fibroin gene (8) and dopa decarboxylase (DDC) gene (9). In *Bombyx mori*, an important economic insect as well as a model lepidopteran insect, POUM was reported to regulate the transcription of sericin-1 gene (7) and to regulate the expression of *B. mori* diapause hormone and pheromone biosynthesis activating neuropeptide (DH-PBAN) gene during metamorphosis (10). In our previous study, BmPOUM2 was shown to regulate the expression of wing disc cuticle protein gene (*BmWCP4*) during metamorphosis of *B. mori*. Down regulation of *BmPOUM2* expression resulted in failure to complete metamorphosis accompanied with wing disc growth and differentiation arrest and failure for larvae to spin cocoons (11,12). Study of the regulatory mechanism of *BmPOUM2* would be critical to explain wing development.

G-quadruplex (G4) and i-motif have been identified as DNA secondary structures. The G4 structure is formed by stacked G-tetrads, which are square co-planar arrays of four guanine bases and stabilized by monovalent cation such as K^+ or Na^+ (13–16). G4 has been found to regulate multiple biological processes such as replication, transcription and telomeres protection (17,18). More than 700 orthologous genes were found to have the probability to form

*To whom correspondence should be addressed. Tel: +86 20 85215291; Fax: +86 20 85215291; Email: qlfeng@scnu.edu.cn

G4 in human, chimpanzee, mouse and rat, suggesting the widespread regulatory influence of G4 motifs (19). In human, the regulatory regions of >40% genes were predicted to have G4 structure and to be involved in the gene regulation at the level of transcription (20). Stable G4 structure identified in the promoter region of the *WNT1* gene suppressed its expression and its downstream signaling pathways, resulting in the inhibition of *Wnt1* mediated migration and invasion activities of cancer cells (21). Similar effect was also demonstrated in the promoter of *c-MYC* (22,23). In contrast, G4 located the upstream transcriptional starting site (TSS) of *col2al*, *fdz5* and *nog3* act as an enhancer for transcription (24).

The complementary strand of G-rich sequence is unsurprisingly C-rich and it can form another DNA secondary structure called i-motif. I-motif is a four-stranded structure formed by intercalated hemiprotonated cytosine–cytosine ($C\text{-}C^+$) base pairs under acidic condition (25) or at neutral pH by molecular crowding of the cosolutes (26) or in high pressure (27). Unlike G4 structure, studies on the i-motif structure and its function are limited. I-motif structure has been found in the promoter of genes *BCL2* (28), *Th* (29) and *KRAS* (30) and has been identified as a recognition binding site for transcription factors to activate gene transcription (31).

In this study, we identified G4 and i-motif secondary structures in the promoter region of the transcription factor *BmPOUM2* and demonstrated for the first time that a nuclear transcription factor protein, BmILF, bound to the i-motif structure, regulating the transcription of *BmPOUM2* in *B. mori*.

MATERIALS AND METHODS

Insect and cell line

Bombyx mori strain Dazao was provided by the Research and Development Center of the Sericultural Research Institute of the Academy of Agricultural Sciences of Guangdong Province, China. Larvae were raised on fresh mulberry leaves at 26°C under a 14/10 h light/dark photoperiod.

A *B. mori* cell line DZNU-*Bm*-12 (*Bm*12) originally developed from ovarian tissues (32) was maintained at 28°C in Grace medium (Invitrogen, USA) supplemented with 10% FBS.

Cloning of the *BmPOUM2* promoter

Genomic DNA was prepared from fifth instar larvae of *B. mori* using the phenol extraction method (33). The primers used for amplifying the promoter of *BmPOUM2* from genomic DNA were: Forward primer: 5′-TGATGACACGACACAATAACC-3′; Reverse primer: 5′-GGCAGGCTCAGCACTACG-3′. Genomic DNA was denatured at 94°C for 3 min, followed by a 30-cycle reaction (94°C for 30 s; 55°C for 30 s; 72°C for 2 min). The primers used for amplifying the promoter of *BmPOUM2* from 1638-pMD-18T vector were: Forward primer: 5′-GTTTATTAAATTGGGGCCGTGA-3′; Reverse primer: 5′-ACGGCAGCACGGAGCGA-3′. Template DNA was denatured at 94°C for 3 min, followed by 27 cycles of 94°C for 30 s, 55°C for 30 s and 72°C for 30 s for each cycle. In this study, different doses of dimethyl sulphoxide (DMSO) were added into the PCR reaction mix or the PCR reaction conditions were adjusted to obtain the target products. PCR products were separated on 1% agarose gels for verification.

Circular dichroism (CD) analysis

CD experiments were performed with a J-815 CD spectrometer (Jasco International, USA). All spectra were collected with a wavelength range from 220 to 350 nm, 1 nm step width and 1 s response time. The CD spectra are representations of three averaged scans of the same sample taken at room temperature and are baseline corrected for signal contributions due to the buffer. DNA oligonucleotide sequences (5 μM) used for analyzing G4 structure were heated at 95°C for 10 min in 50 mM Tris buffer (pH 7.5) with or without 100 mM KCl, and slowly cooled to room temperature over 4 h period. DNA oligonucleotide sequences (5 μM) used for analyzing i-motif structure were heated at 95°C for 10 min in 50 mM Tris–acetate buffer at different pHs, and slowly cooled to room temperature over 4 h period. For the incubation with TMPyP2 or TMPyP4 (Frontier Scientific, USA), TMPyP2 or TMPyP4 (dissolved in ddH_2O) was added into the annealed mix at a final concentration of 25, 50, 100 μM and incubated overnight at 4°C. The CD melting curves were obtained at 286 nm between 20 and 95°C with a heating rate of 1°C/min. The synthesized ssDNA oligonucleotides that contained the i-motif region (5 μM) were heated at 95°C for 10 min in 50 mM Tris-acetate buffer at pH 4.1, and slowly cooled to room temperature over 4 h period to allow i-motif structure to form. TMPyP2 or TMPyP4 was then added into the solution at a final concentration of 25 μM and incubated overnight at 4°C, followed by CD analysis.

Luciferase assay for promoter activity analysis

*Bm*12 cells were inoculated into culture media in 12- or 24-well culture plates (Corning, USA) and cultured for 12 h. Cell transfection and co-transfection were conducted when the cell density achieved 80%. To normalize the firefly luciferase activity, the renilla luciferase vector pRL-SV40 (Promega, USA) was co-transfected with each of the pGL3-derived vectors. For transfection, a mixture of 50 μl containing 200 ng of the wild-type or mutated 1360-BmPOUM2-P-pGL3 reporter plasmid DNA, 50 ng internal control plasmid and 2 μl Fugene HD transfection reagent (Promega, USA) in the Opti-MEM Reduced Serum Medium (Life Technologies, USA) was added to *Bm*12 cells in Grace medium with or without TMPyP2 or TMPyP4. The cells were cultured for 48 h at 28°C before the promoter activity assay.

For co-transfection, a mixture of 50 μl containing 200 ng of the wild-type or mutated 1360-BmPOUM2-P-pGL3 reporter plasmid DNA, 1 μg EGFP or BmILF-EGFP plasmid DNA, 200 ng internal control plasmid (pRL-SV40 vector) and 6 μl Fugene HD transfection reagent in the Opti-MEM Reduced Serum Medium was added to the cells in Grace medium. The cells were cultured for 48 h at 28°C before the promoter activity assay.

For the anti-sense oligonucleotides (ASOs) disturbing assays, a mixture of 50 μl containing 200 ng of

the wild-type or mutated 1360-BmPOUM2-P-pGL3 reporter plasmid DNA, 2 μl ASOs (100 μM), 300 ng internal control plasmid (pRL-SV40 vector) and 6 μl Fugene HD transfection reagent in the Opti-MEM Reduced Serum Medium was added to the cells in Grace medium in each well. The cells were cultured for 48 h at 28°C before the determination of promoter activity. The oligonucleotides used in the ASOs disturbing assays were: wild-type ASO-i-motif: 5′-GGGGCGCGAGGG-3′; mutant ASO-i-motif: 5′-GATGCGCGAGTG-3′; wild-type ASO-G4: 5′-CCCTCGCGCCCC-3′; mutant ASO-G4: 5′-CACTCGCGCATC-3′; non-specific-ssDNA: 5′-ATCGATCGATCG-3′; F-ASO-β-actin: 5′-TTTCGCGTGTAT-3′; R-ASO-β-actin: 5′-ATACACGCGAAA-3′.

The measurement of luciferase activity was carried out as described previously (11). Briefly, the transfected cells were washed twice with PBS and lysed in 150 μl Passive Lysis Buffer (Promega, USA). Samples were centrifuged at 1000 g for 5 min at room temperature. The supernatant was used to analyze the luciferase activity using the Dual-Luciferase Assay System according to the manufacturer's protocol with a luminometer (IBA7300, Veritas, Turner Biosystems). The luciferase activity was normalized to the renilla luciferase activity. 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.

Protein pull-down and liquid chromatography-tandem mass spectrometry (LC-MS/MS) analysis

Protein pull-down assays referenced the previous study (31). Nuclear proteins of *Bm*12 cells were extracted according to the instructions of NE-PER Nuclear and Cytoplasmic Extraction Reagents (Thermo Scientific, USA). The protein concentration was determined using the BCA Protein Assay Reagent Kit (Thermo Scientific, USA). DNA oligonucleotide sequences of G4 structure were heated at 95°C for 10 min in 50 mM Tris buffer at pH 7.5 and slowly cooled to room temperature over 4 h. DNA oligonucleotide sequences of i-motif structure were heated at 95°C for 10 min in 50 mM Tris-acetate buffer at pH 4.1 and slowly cooled to room temperature over 4 h. Single-stranded biotinylated DNA oligonucleotide (20 μg) was incubated with 100 μg streptavidin-coated Dynabeads (Life Technologies, USA) in 400 μl binding buffer (10 mM Tris, pH 7.5, 1 mM EDTA, 1 M NaCl, 0.003% NP40) for 30 min at room temperature with constant and slow rotation. After twice washing with binding buffer, the immobilized DNA was incubated for 30 min in 400 μl blocking buffer (2.5 mg/ml BSA, 10 mM HEPES, pH 7.6, 100 mM potassium glutamate, 2.5 mM DTT, 10 mM magnesium acetate, 5 mM EGTA, 3.5% glycerol with 0.003% NP40 and 5 mg/ml polyvinylpyrrolidone). Then 100 μg nuclear proteins was incubated with the immobilized DNA for 4 h at 4°C in 400 μl protein binding buffer (10 mM HEPES, pH 7.6, 100 mM potassium glutamate, 80 mM KCl, 2.5 mM DTT, 10 mM magnesium acetate, 5 mM EGTA, 3.5% glycerol with 0.001% NP40 and 1 μg non-specific DNA) with constant and slow rotation. After 4 h, the DNA/protein complexes were washed three times with 400 μl washing buffer (10 mM HEPES, pH 7.6, 100 mM potassium glutamate, 2.5 mM DTT, 10 mM magnesium acetate, 5 mM EGTA, 3.5% glycerol, 0.5 mg/ml BSA, 0.05% NP40). Then the proteins bound to the DNA were eluted in 20 μl SDS-PAGE sample buffer (50 mM Tris, 100 mM DTT, 2% SDS, 0.1% Bromophenol blue, 10% glycerol). The eluted proteins were subjected to 12% SDS-PAGE. The gels were stained with Coomassie Brilliant Blue R-250 for more than 4 h and distained with distaining solution (10% acetic acid, 5% ethanol, 85% water). The differentiated protein bands were excised and analyzed for protein identification in Huijun Biotechnology (Guangzhou, China) using liquid chromatography-tandem mass spectrometry (LC-MS/MS).

Expression and purification of recombinant BmILF protein

Total RNA extraction from *Bm*12 cells and reverse transcription were conducted using the same method as described previously (11). *BmILF* open reading frame (ORF) fragment was amplified using cDNA cloned from *Bm*12 cells with forward primer: 5′-GAATTCATGGTGCGCGGCATC-3′ and reverse primer: 5′-AAGCTTTTACACT GAATCTTTGTCGTCGGT-3′. Obtained DNA fragment was then sub-cloned into the pET-28a vector in fusion with a 6 × His tag between EcoR I and Hind III restriction enzyme sites, generating a BmILF-pET-28a recombinant expression vector.

The recombinant BmILF protein was expressed in *Escherichia coli* cells (BL21), which were grown in Luria-Bertani (LB) medium containing 50 μg/ml kanamycin at 18°C with 12 h induction of 0.1 mM Isopropyl-β-D-thiogalactopyranoside (IPTG). The cells were collected by centrifugation and re-suspended in the binding buffer (0.5 M NaCl, 20 mM Tris–HCl, 5 mM Imidazole, pH 7.9 and 1 mM PMSF). The suspension was centrifuged at 10 000 g at 4°C for 5 min after being lysed by sonication, and then purified with Ni-chelating affinity chromatography using the His-Bind® Kit according to the manufacturer's protocol (Novagen, USA). The protein concentration was determined using the BCA Protein Assay Reagent Kit (Thermo Scientific, USA)

Electrophoretic mobility shift assay (EMSA) for protein-DNA binding

EMSA was conducted using the Light Shift Chemiluminescent EMSA Kit (Thermo Scientific, USA). The wild-type and mutant oligonucleotides were labeled with biotin at the 5′ end and heated at 95°C for 10 min in 50 mM Tris-acetate buffer at pH 4.1 and slowly cooled to room temperature. The biotin labeled oligonucleotides at the 5′ end was synthesized in Invitrogen (Shanghai, China).

Binding reactions were performed according to the instruction of the EMSA Kit. Briefly, reactions were conducted in a 20 μl mixture containing (1 × binding buffer, 2.5% glycerol, 0.05% NP-40, 5 mM $MgCl_2$, 4 mM EDTA, recombinant protein BmILF and 20 fmol of a biotinylated end-labeled probe) at room temperature for 20 min. For the competition assay, cold probes (un-biotinylated) were added to the binding reaction. The samples were then separated in a 6% or 12% polyacrylamide gel on

the ice at 100 V for 1.5 or 3 h. After electrophoresis, the gel was blotted onto a Nylonpositively charged membrane (Amersham Biosciences, USA). The membranes were then developed using the Light Shift Chemiluminescent EMSA Kit according to the manufacturer's protocol. The oligonucleotide probes used were showed as followed: Wild-type probe: 5′-TTGTTGCCCCGCCCCTCGGCCCCCT CGCGCCCCGCAC TGG-3′; Mutant probe (the mutant sites were underlined): 5′-TTGTTGC ACTGCTACTCG GCTCACTCGCGCATCGCACTGG-3′.

Chromatin Immunoprecipitation (ChIP) assay

ChIP assay was conducted to detect the binding of BmILF with the i-motif of the *BmPOUM2* promoter in *Bm*12 cells. Cells were crosslinked with 1% formaldehyde for 10 min at room temperature after being transfected with BmILF-EGFP or EGFP (control) plasmids for 48 h. Glycine was added to terminate the fixation and the cells were washed twice with one media volume of ice-cold PBS. Cells in 1 ml of ice-cold PBS and 10 μl Halt Cocktail were collected by scraping. The cells were centrifuged at 3000 g for 5 min and the cell pellet was broken up with extraction buffer containing protease/phosphatase inhibitors. Nuclei were collected by centrifuged at 9000 g for 3 min and digested by MNase. Digested chromatin was obtained by sonicating on ice with several pulses to break nuclear membrane and incubated for 20 s on ice between pulses and then centrifuged at 9000 g for 5 min. Immunoprecipitation experiment was performed following the manufacturer's instruction of Pierce™ Magnetic ChIP Kit (Thermo Scientific, USA). Ten micrograms of either rabbit anti-EGFP antibody or normal rabbit IgG (Thermo Scientific, USA) were used and IP reactions were incubated overnight at 4°C with constant mixing. DNA/protein/antibody complex was purified by incubating with ChIP Grade Protein A/G Magnetic Beads for 2 h at 4°C with mixing. Immunoprecipitated genomic DNA fragments were amplified by RT-PCR and qRT-PCR with primers: forward primer: 5′-GCGAGACGTACAGACGCGAG-3′; reverse primer: 5′-GCCCGCTAATGACGGTGTAA-3′. The length of the product is 86 bp. The specificity of primers was examined using Primer-blast (http://www.ncbi.nlm.nih.gov/tools/primer-blast/). For RT-PCR, template DNA was denatured at 96°C for 3 min, followed by 26 cycles of 96°C for 30 s, 55°C for 30 s and 72°C for 10 s for each cycle. For qRT-PCR, the SYBR Green Kit was used according to the manufacturer's instruction (TaKaRa, Japan). The qRT-PCR was performed at the following conditions: SYBR Premix Ex Taq (2×): 10 μl in 20 μl reaction volume; the primers concentrations: 0.4 μl (10 μM); the immunoprecipitated DNA samples: 4 μl. The mixtures were incubated at 95°C for 10 s, then 40 cycles at 95°C for 5 s, 60°C for 31 s using ABI7300 fluorescence quantitative PCR system. The enrichment of promoter sequence in immunoprecipitated DNA samples was normalized with DNA present in the 10% input material analyzed using the $2^{-\triangle\triangle Ct}$ method (34). Data is means ± SEM ($N = 3$). ***$P < 0.001$ (Student's *t*-test). The PCR products of the enriched promoter were sequenced.

RNA interference (RNAi)

For RNAi of BmILF in *Bm*12 cell line, a 458 bp unique fragment from 198–655 nts in the ORF of BmILF was chosen as a template for synthesizing gene-specific dsRNA. BmILF dsRNA was synthesized from the linearized template by using the T7 RiboMAXTM Express RNAi System (Promega, USA). DsRNA (4 μg) was transfected into *Bm*12 cells in Grace medium with 6 μl Fugene HD transfection reagent in the Opti-MEM Reduced Serum Medium. The cells were collected 48 and 72 h after transfection. Total RNA extraction and reverse transcription were conducted as above mentioned (11). RT-PCR was performed using specific primers designed based on the *BmILF* cDNA to detect the knock-down efficiency. The sequence of primers were: forward primer: 5′-ATGGTGCGCGGCATC-3′, reverse primer: 5′-TTACACTGAATCTTTGTCG TCGGT-3′. Template DNA was denatured at 94°C for 3 min, followed by 29 cycles of 94°C for 30 s, 55°C for 30 s and 72°C for 30 s for each cycle. To detect the effect of BmILF on *BmPOUM2* promoter activity when *BmILF* was knocked down, a 50 μl mixture containing 4 μg *BmILF* dsRNA, 200 ng of the wild-type or mutated 1360-BmPOUM2-P-pGL3 reporter plasmid DNA, 200 ng internal control plasmid (pRL-SV40 vector, Promega) and 6 μl Fugene HD transfection reagent (Promega, USA) in the Opti-MEM Reduced Serum Medium (Life Technologies, USA) was added to cells in Grace medium. The cells were cultured for 72 h at 28°C before the promoter activity assay.

Genomic search for silkworm G-quadruplex and i-motif

BmPOUM2 orthologues in the six insect species were identified by blastp against NCBI nr database (http://www.ncbi.nlm.nih.gov/) and their upstream ∼500 bp were retrieved from their genomic sequences in NCBI GenBank (http://www.ncbi.nlm.nih.gov/). The accession numbers of the orthologues are XP_013185201.1 (*Amyelois transitella*), EHJ67329.1 (*Danaus plexippus*), XP_013178065.1 (*Papilio xuthus*), ACN43331.1 (*Tribolium castaneum*) XP_393686.2 (*Apis mellifera*), NP_001163355.2 (*Drosophila melanogaster*). Sequences were aligned using ClustalW package in Bioedit (Version 7.2.5) and manually annotated.

Genomic sequences, geneset and related annotation data of the silkworm were obtained from SilkDB (http://silkworm.swu.edu.cn/). Motifs of G4 (G3+N1–7G3+N1–7G3+N1–7G3+) and the corresponding i-motif were searched in the silkworm genome, respectively, using the program Quadparser (35). These motifs were then localized to specific genomic annotated regions with ≥ 1 bp overlapping cutoff.

RESULTS

Identification of the secondary structures in the promoter of *BmPOUM2*

To clone the promoter sequence of the transcription factor gene *BmPOUM2*, PCR amplification was performed using the primers designed based on the *B. mori* genome sequence deposited in SilkDB with genomic DNA isolated from *B.*

mori larvae as templates. However, all trials for cloning the target sequence were failed under the general PCR condition. When 5% or 8% DMSO was added to the PCR reaction mixture under the same reaction conditions, the target promoter sequence (1638 bp) was obtained (Figure 1A). This result implied that the target promoter region may contain DNA secondary structures that block the Taq DNA polymerase binding with the DNA (36,37).

For further investigation, the 1638 bp promoter sequence was cloned into pMD-18T vector, generating a 1638-pMD-18T recombinant plasmid. Because the G4 and i-motif are always found in the GC-rich region of DNA, the PCR primers were designed to amplify the GC-rich region only in the subsequent DMSO dose-dependent PCR. When the concentration of DMSO was increased higher than 3%, the target sequence (384 bp) started to be amplified (Figure 1B). When the denaturation temperature of PCR reaction was risen to 96°C or 97°C, the target sequence was also amplified (Figure 1C). When the denaturation temperature was maintained at 95°C but the denaturation time was extended to 60 s, the target sequence could also be amplified (Figure 1D). All these results implied that the *BmPOUM2* promoter may contain secondary structures that can be deformed either by chemical (DMSO) or physical (higher denaturation temperature or longer denaturation time) factors.

Location of the secondary structures

Sequence analysis of the cloned promoter of *BmPOUM2* indicated that it has a GC-rich region between –237 and +1 nucleotides (nts) (Figure 2A). To detect the accurate position of the secondary structures, the G-rich region of the forward strand of the promoter was truncated into three parts: –78 to +1 nts; –79 to –157 nts and –158 to –237 nts and subjected to CD analysis. The results indicated that the –78 to +1 and –158 to –237 regions had maximum and minimum absorption peaks at 280 and 250 nm, respectively, whereas the –79 to –157 region had maximum and minimum absorption peaks at 270 and 240 nm, respectively, and the maximum peak at 270 nm was enhanced when 100 mM K^+, which stabilizes the G4 structure (38), was added to the DNA sample (Figure 2B). This CD spectrum change was well consistent with the reported pattern of absorption peaks of reported G4 structure (24,29,39).

QGRS Mapper (http://bioinformatics.ramapo.edu/QGRS/) is a program that detects the composition and distribution of putative quadruplex forming G-rich sequences (QGRS) in nucleotide sequences (40). Analysis on the –79 to –157 nts region with QGRS Mapper revealed that the sequence of –94 to –120 region (GGGGCGCGAGGGGGCCGAGGGGCGGGG) has a higher G-score than many reported G4 sequences (39,41–48) (Figure 2C). The forward (F) and reverse (R) single strand DNA (ssDNA) of –94 to –120 nts were examined by CD for G4 and i-motif structures, respectively. For the F-strand, maximum and minimum peaks were observed at 265 and 240 nm, respectively, and the peak at 265 nm was increased when 100 mM K^+ was added (Figure 2D), showing a typical feature of G4 structure (24,29,39). When the guanine at –95, –97, –100, –101, –108, –110, –118 and –119 sites were mutated to T, A, A, T, A, T, T and A, respectively, the maximum and minimum absorption peaks were shifted to longer wavelengths and the K^+-mediated enhancement of the maximum intensity was suppressed (Figure 2D). For the R-strand, maximum and minimum peaks were observed at 290 and 260 nm, respectively, when the ssDNA solution was kept in acidic condition (pH < 6.0). However, the maximum and minimum peaks were shifted to 280 and 250 nm, respectively, at pH ≥6.6 (Figure 2E and Supplementary Figure S1), which is nearly identical to previous analyses of i-motif structures by CD spectroscopy (29,49,50). When the cytosine at –95, –97, –100, –101, –108, –110, –118 and –119 sites were mutated to A, T, T, A, T, A, A and T, respectively, the maximum and minimum absorption peaks were shifted to shorter wavelengths and the H^+-induced enhancement of the maximum intensity was blocked (Figure 2E).

All these results of PCR, bioinformatic prediction and CD analysis suggested that a G4 and an i-motif structure were present in the F- and R-strand, respectively, of the –94 to –120 nts region of the *BmPOUM2* promoter.

Effects of the i-motif structure on the transcription

To test whether or not the secondary structures are involved in the regulation of *BmPOUM2* transcription, luciferase expression activity assay was performed. The wild-type or mutated promoter sequence (+72 to –1360 nts) was cloned into an expression vector (pGL3) to control the luciferase expression and the recombinant vector was used to transfect *Bm*12 cells. The results indicated that the mutated structure significantly decreased the activity (∼75%) of the promoter as compared to the wild-type structure (Figure 3A), suggesting that the secondary structure, either G4 or i-motif, or both, is necessary for the promoter activity.

Addition of excess amount of ASOs complementary to a secondary structure sequence could inhibit its formation, allowing formation of normal dsDNA and resulting in inhibition or activation of transcription (24,51). This assay is not involved in any base mutation and can maintain one (either G4 or i-motif) structure while suppress the other, allowing to determine which structure contributes the regulation function (Figure 3B). The results of the ASO assay indicated that when the wild-type ssDNA that is complementary to the i-motif sequence (ASO-i-motif) was used to treat the cells, the promoter activity was significantly decreased (Figure 3C), implying the i-motif was suppressed by the excess amount of ASO-i-motif. The mutated ASO-i-motif did not have this inhibitory effect. When either wild-type or mutated ssDNA that are complementary to the G4 structure sequence (ASO-G4) was added, the promoter activity did not significantly change (Figure 3C). As negative control, complementary ssDNA fragments of the *β-actin* promoter (ASO-*β-actin*) did not change the activity because it does not contain any secondary structures.

The results of ASO assay indicated that only the i-motif structure participated in the regulation of *BmPOUM2*. Therefore, the regulatory function of i-motif structure was only focused in the subsequent researches. The porphyrin compound TMPyP4 (Figure 4A) has been reported to be able to bind with the G4 and i-motif structures (52,53). The TMPyP4 isoform, TMPyP2, cannot bind with G4 and i-

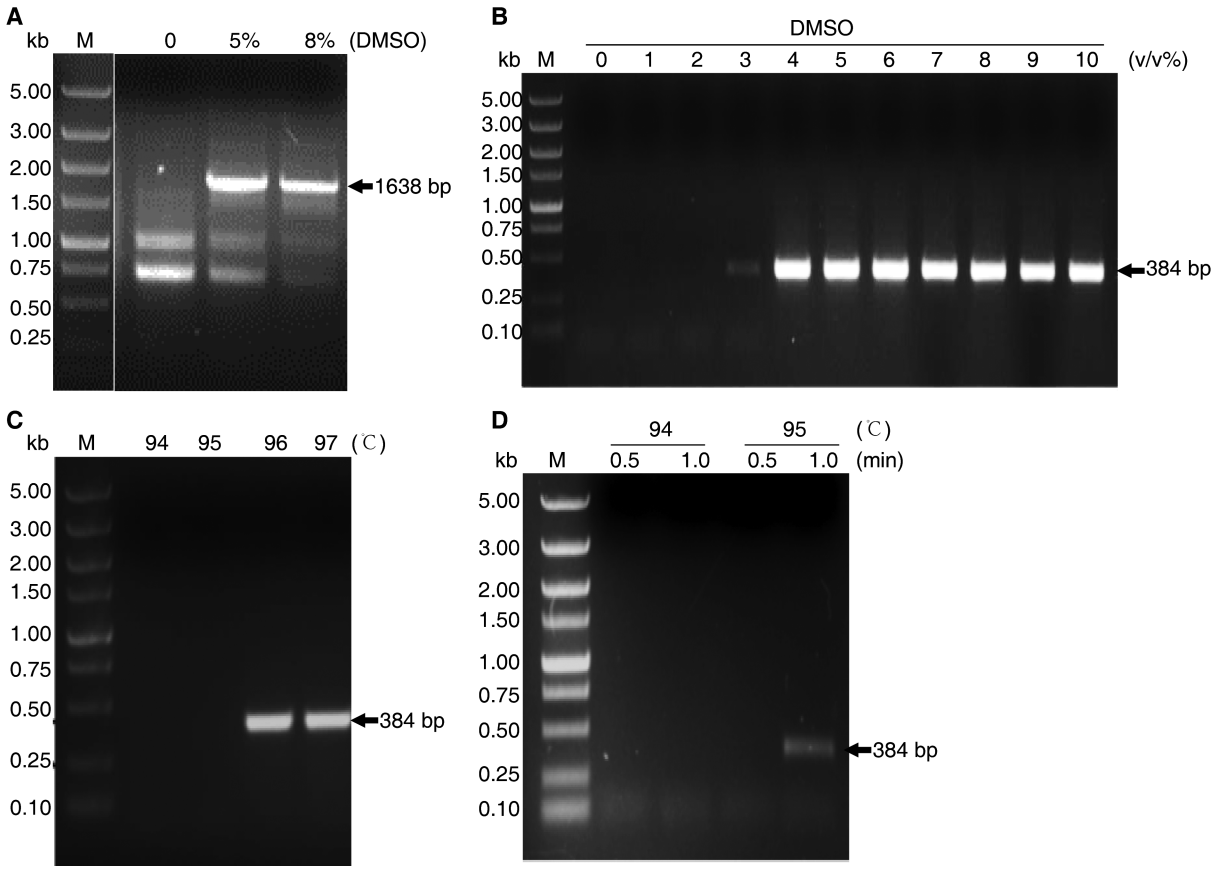

Figure 1. PCR amplification of the promoter sequence of *BmPOUM2* at different reaction conditions. (**A**) PCR reaction in the presence of 0, 5 and 8% DMSO. The faster migrating small bands are the non-target PCR products, non-specifically amplified. (**B**) The DMSO dose-dependent PCR assay with 0–10% (v/v) of DMSO in the PCR reaction mix. (**C**) The effect of denaturation temperatures on the amplification of PCR product without DMSO. (**D**) The effect of denaturation time on the amplification of PCR product. The denaturation temperatures kept at 94 or 95°C and the denaturation time were kept for 0.5 or 1.0 min. The general PCR conditions are described in Materials and Methods. The arrow shows the targeted PCR product.

motif structures and can be used as negative control (29,54). To determine whether the *BmPOUM2* i-motif can interact with TMPyP4, CD analysis was performed. The results indicated that TMPyP4, compared to TMPyP2, could dramatically deform the i-motif structure when the concentration reached 25 μM. When the concentration of TMPyP4 exceeded 50 μM the i-motif almost disappeared (Figure 4B). In addition, the CD analysis of melting temperature indicated that the TMPyP4 could dramatically decrease the melting temperature (T_m from 62 to 46°C) of the ssDNA fragment that contained the i-motif structure (Figure 4C). TMPyP2 did not have this effect. These results implied that the *BmPOUM2* i-motif structure can be deformed by TMPyP4.

If this promoter sequence can form i-motif structure in cells, the promoter activity of *BmPOUM2* should be changed when the cells were treated with TMPyP4. In the promoter activity assay, TMPyP4 or TMPyP2 was used to treat the cells when the recombinant vectors were transfected in cells. The results showed that comparing to negative control TMPyP2, TMPyP4 significantly suppressed the wild-type promoter activity (Figure 4D). However, when the nucleotides at -95, -97, -100, -101, -108, -110, -118 and -119 sites were mutated to A, T, T, A, T, A, A and T, respectively, the inhibition effect of TMPyP4 disappeared (Figure 4D), suggesting that TMPyP4 interacted with the wild-type structure of i-motif and impacted the transcription function of the promoter.

These results demonstrated that when the i-motif structure was deformed by site mutation, ASOs or TMPyP4, the promoter activity of *BmPOUM2* was decreased, revealing that the i-motif structure is involved in the regulation of the *BmPOUM2* transcription.

Identification of nuclear protein that binds to the i-motif structure

To identify the nuclear protein that binds to the i-motif structure and regulates the transcription of *BmPOUM2*, pull-down experiments were conducted. The nucleotide fragment of the i-motif structure was labeled with biotin and then linked to the streptavidin-coated beads. The complex was then incubated with the nuclear proteins isolated from *Bm*12 cells. The proteins that bound to the biotin-labeled i-motif probe were separated on a SDS-PAGE gel and the differentiated protein bands were excised and subjected to LC–MS/MS analysis. The results showed that there was one differentiated protein band for the i-motif probe, compared to the mutated probe (which could not

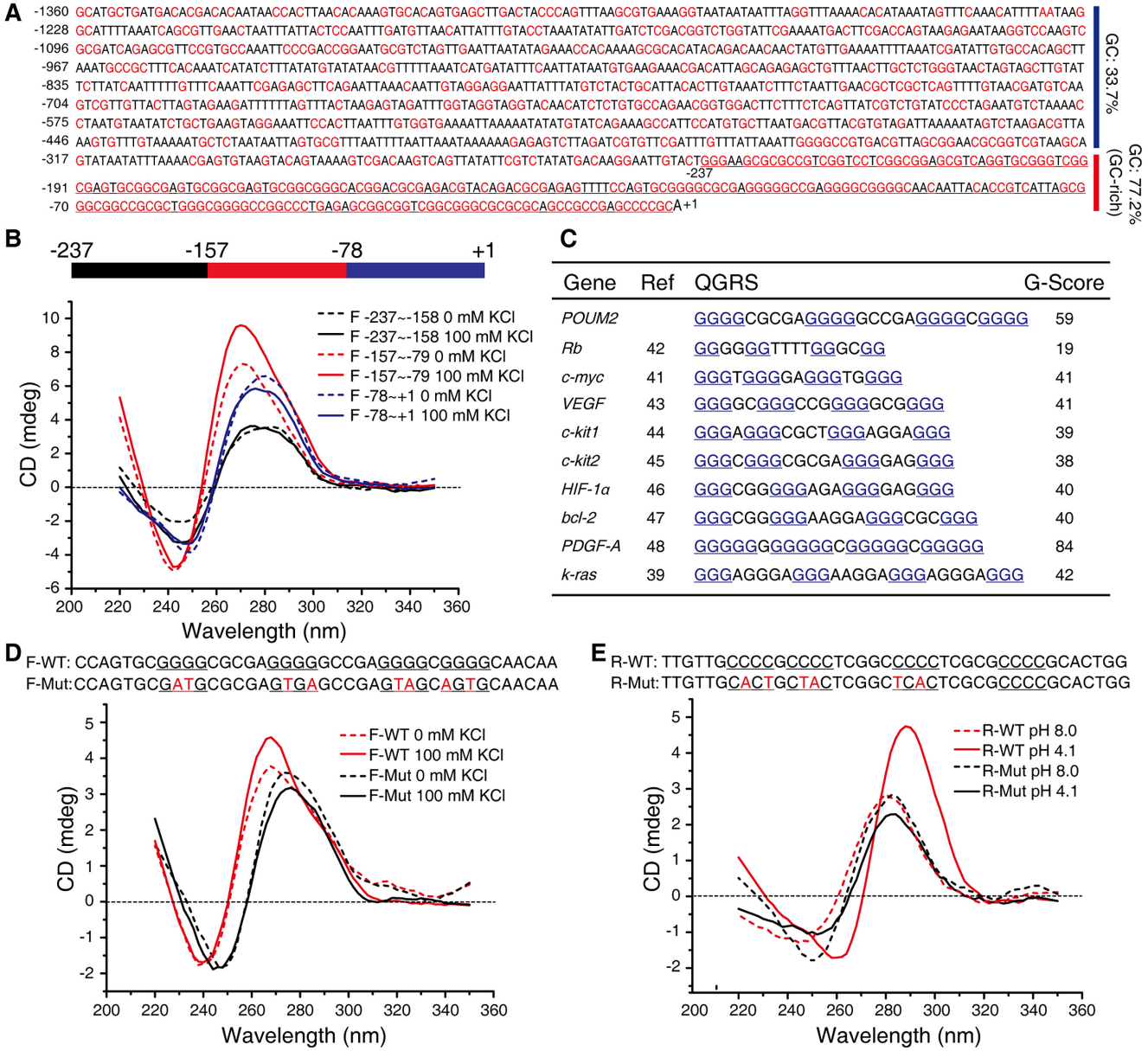

Figure 2. Identification of the G4 and i-motif structures in the promoter region of the *BmPOUM2* gene. (**A**) The nucleotide sequence of the promoter region of the *BmPOUM2* gene. The G and C bases are shown in red. (**B**) CD analysis of the +1 to –237 nts region of the promoter truncated into three fragments with or without 100 mM KCl. (**C**) Comparison of the G4 region of the *BmPOUM2* promoter and some reported G4-containing sequences predicted by the QGRS Mapper. (**D**) CD analysis of the forward ssDNA of the –88 to –127 nts region of the promoter in absence or presence of 100 mM KCl. The mutated nucleotides are shown in the mutant sequence. F-WT: forward wild type strand; F-Mut: forward mutated strand. (**E**) CD analysis of the reverse ssDNA of the –88 to –127 nts region of the promoter at pH 4.1 and 8.0. The mutated nucleotides are shown in the mutant sequence. R-WT: reverse wild type strand; R-Mut: reverse mutated strand.

form i-motif structure). The size of the protein band for the i-motif probe was approximate 41.5 kDa (Figure 5A). While the identification of the G4 binding proteins still is ongoing, a homolog of the human interleukin enhancer binding factor (ILF, GenBank accession no.: NM_001046918.1) (55) was identified from the candidate i-motif binding proteins (Supplementary Table S1). To determine whether this protein binds to the i-motif, the ORF of the *B. mori* ILF (BmILF, GenBank accession no.: KY082711) was cloned. A recombinant protein was expressed in *E. coli* and purified for EMSA. The EMSA assay showed that the recombinant BmILF specifically bound to the labeled i-motif probe and the binding was enhanced with the increase of the protein concentration and competed off by the unlabeled probe (Figure 5B). The BmILF did not bind to the mutated probe, which could not form the i-motif. To further demonstrate that the BmILF was binding to the i-motif secondary structure rather than to the linear ssDNA, the concentration of native-PAGE was increased from 6% to 12%. In 12% native-PAGE the patterns of the i-motif and linear ssDNA were distinguished (Figure 5C). In the absence of BmILF, the DNA probe at pH 4.1/6.0 was separated into two different

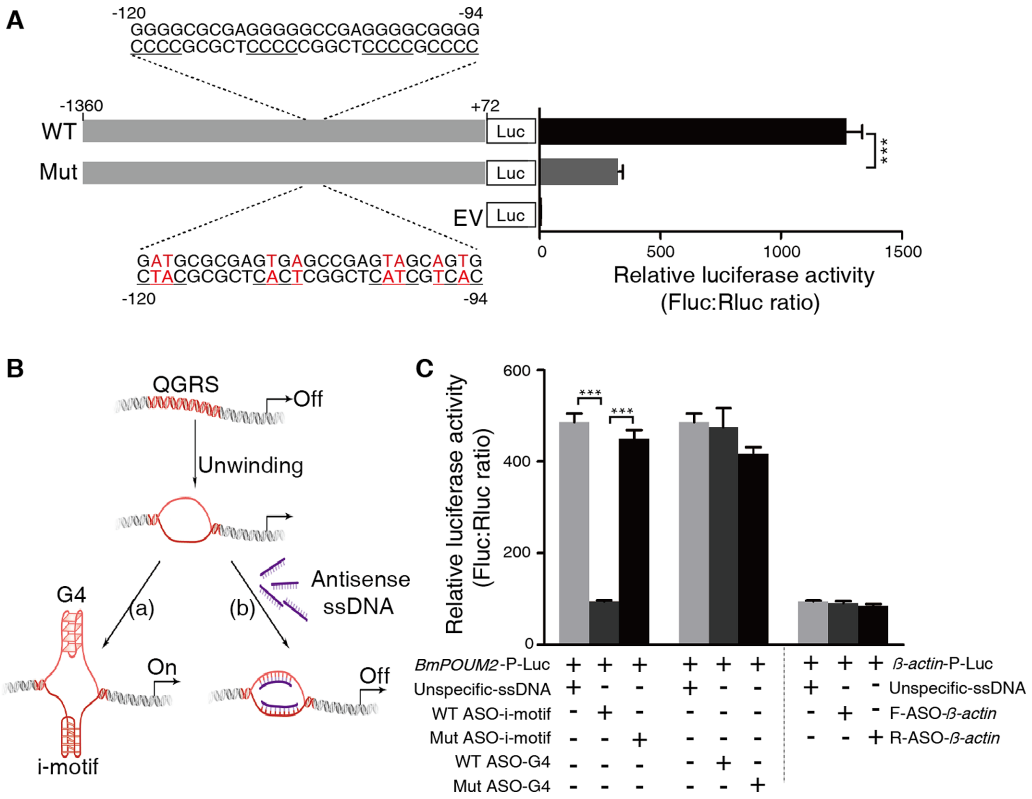

Figure 3. Effects of mutation and ASOs on the promoter activity of BmPOUM2. (**A**) Determination of the activity of the *BmPOUM2* wild-type or mutated promoter by the luciferase assay. WT, wild-type; Mut, mutant; EV, empty vector. (**B**) Descriptive diagram of the principle of ASO assay. In normal conditions, the dsDNA was unwound to ssDNA and the G-rich and C-rich ssDNA fold to form G4 and i-motif structures, respectively (a); when the excessive ASOs of the G4 or i-motif were added, the G4 or i-motif would be inhibited (b). (**C**) Effect of ASOs interfering assay on the *BmPOUM2* promoter activity. Non-specific-ssDNA: non-specific single-stranded DNA was used as a control. WT-ASO-i-motif and Mut-ASO-i-motif: wild-type and mutated complementary anti-sense oligonucleotides of i-motif. WT-ASO-G4 and Mut-ASO-G4: wild-type and mutated complementary anti-sense oligonucleotides of G4. F-ASO-*Bmβ-actin* and R-ASO-*Bmβ-actin*: complementary anti-sense oligonucleotides of reverse and forward strand. Data is means ± SEM ($n = 3$). ***$P < 0.001$ (Student's t test).

bands, but at pH 8.0, only one band was found (Figure 5C). So, the lower bands were the i-motif structure and the higher bands were linear ssDNAs. When BmILF was added to the reaction system, the i-motif probe band was decreased and a new higher band (BmILF+i-motif probe) was appeared. The linear ssDNA band was not affected (Figure 5C). This result indicated that BmILF did bind to the i-motif structure, rather than to the linear ssDNA.

To further demonstrate whether or not BmILF can bind to the i-motif of the *BmPOUM2* promoter *ex vivo*, ChIP assay was performed. An EGFP tag was fused to BmILF and anti-EGFP tag antibody was used in the ChIP assays (56). The cells were transfected with the recombinant plasmid BmILF-EGFP or EGFP (as a control) (Figure 5D) and then collected for ChIP assay after 48 h post transfection. The results showed that the anti-EGFP antibodies, but not IgG (as a negative control), precipitated i-motif fragment of the *BmPOUM2* promoter in the cells transfected with the BmILF-EGFP expressing plasmid (Figure 5D). The enriched i-motif sequence amplified by PCR was confirmed by sequencing (Figure 5E and F). This experiment demonstrated that BmILF did bind to the i-motif of the *BmPOUM2* promoter *ex vivo*.

The regulatory function of BmILF

To determine whether or not BmILF influences the *BmPOUM2* promoter activity, the BmILF-EGFP expression construct and the pGL3 *BmPOUM2* promoter-luciferase construct were co-transfected into *Bm*12 cells and the reporter activity was quantified after 48 h post transfection (Figure 6A). The *BmPOUM2* promoter-controlled luciferase activity was increased significantly in the cells transfected with BmILF-EGFP, but not in the control cells transfected with EGFP (Figure 6B). Complementary to the over-expression experiment, BmILF RNAi was performed. When the BmILF expression was down-regulated by dsBmILF in *Bm*12 cells (Figure 6C), the activity of the *BmPOUM2* promoter was significantly decreased (Figure 6D).

All the experimental results of the up-regulated and down-regulated BmILF expression suggested that the nuclear protein BmILF bound to the i-motif structure in the *BmPOUM2* promoter and regulated its transcription.

DISCUSSION

G4 structure widely exists in human genome (57,58) and it has been found to exist in the regulatory regions of genes and distribute near the transcription start sites (TSSs)

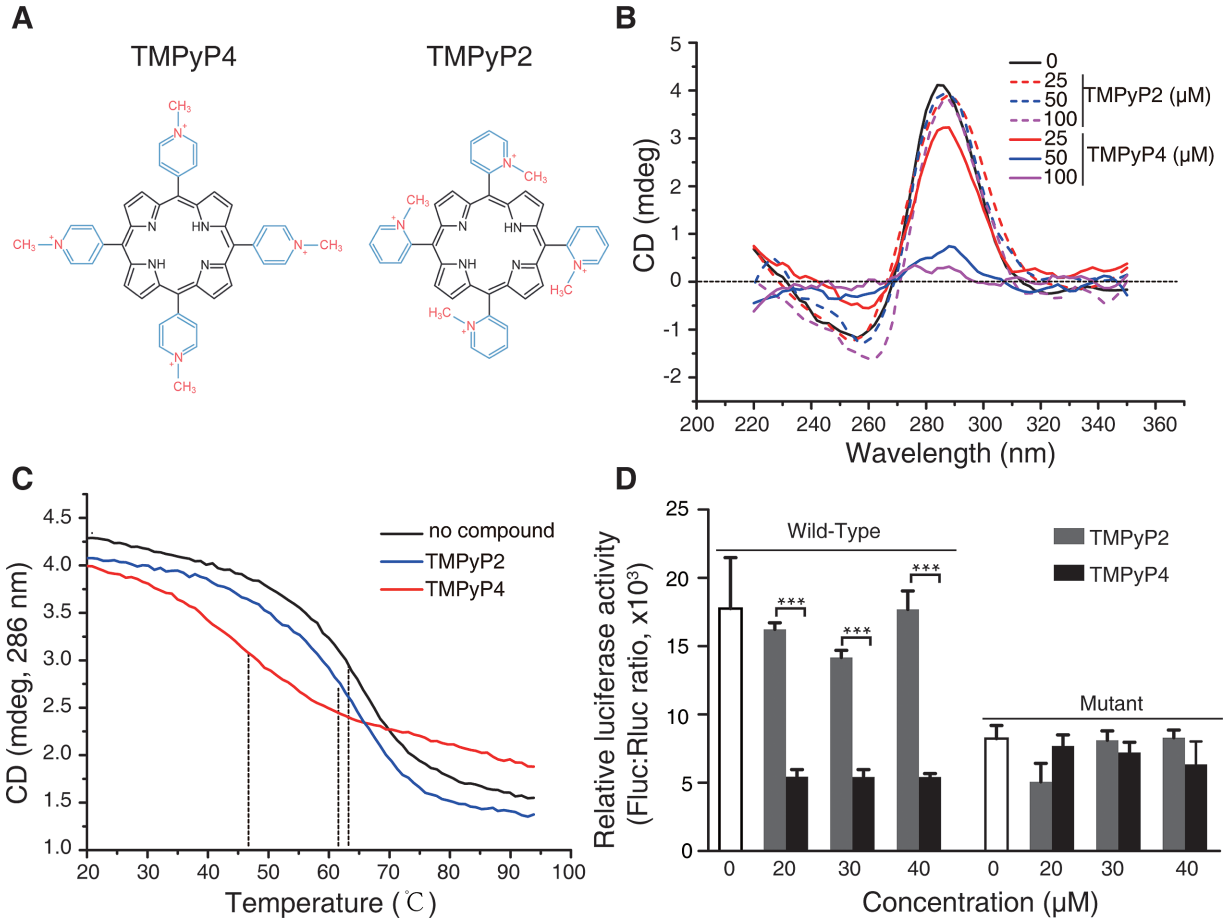

Figure 4. Effects of TMPyP4 and TMPyP2 on the i-motif structure and promoter activity of *BmPOUM2*. (**A**) The structure of the porphyrin compounds TMPyP2 and TMPyP4. (**B**) CD analysis of the i-motif structure in the presence of TMPyP4 or TMPyP2. (**C**) CD analysis of melting temperature of the i-motif structure in the presence of TMPyP4 or TMPyP2. The synthesized ssDNA oligonucleotides that contained the i-motif region (5 μM) were heated at 95°C for 10 min in 50 mM Tris-acetate buffer at pH 4.1 and slowly cooled to room temperature over 4 h period to allow i-motif structure to form. TMPyP2 or TMPyP4 was then added into the solution at a final concentration of 25 μM and incubated overnight at 4°C, followed by CD analysis. The dot lines show the melting temperatures. (**D**) Determination of the promoter activity by the luciferase assay in the presence of TMPyP4 or TMPyP2. Data are means ± SEM ($n = 3$). ***$P < 0.001$ (Student's t test).

(19,20), implying that G4 secondary structure may function as an important regulatory *cis*-element. The C-rich complementary strand of G4 can form i-motif, but relatively less studies on the i-motif function have been reported. In this study, G4 and i-motif structures were identified in the GC-rich regions (between –120 and –94) of the *BmPOUM2* promoter, which can be deformed either by chemical (DMSO) or physical (higher denaturation temperature or longer denaturation time) factors. This is the first report on the secondary structures present in the regulatory region of a transcription factor gene in lepidopteran insects.

To examine whether or not these secondary structures are also present in the POU homologue genes and the silkworm genome, the promoter sequences of POU homologue genes of six species were analyzed. Five POUs promoter showed predicted G4 and i-motif structures (Supplementary Figure S2). In the *B. mori* genome, 3174 G4 and 3104 i-motif structures were found with 357 motifs (147 G4 and 210 i-motif) being in the upstream region of 322 genes (2.27% of total gene set) (Supplementary Table S2). These results imply that DNA secondary structures are widely and commonly distributed in the genome and may play important roles in the regulation of gene transcription in insect.

Studies on the transcriptional roles of G4 and i-motif reveal that these structures play inhibitory or promoting effects on gene transcription. In some cases, they may have repressive effect. For example, in *WNT1*, *c-MYC* (21–23), the G4 structure suppresses the gene expression, probably due to prevention of the indispensable transcription factors from binding to the promoter or due to facilitation of binding of transcription suppressors to the promoter. In other cases, the secondary structures have activation effect, for example, in *col2al*, *fdz5*, *nog3* (24) and muscle specific gene promoter (59), probably due to recruitment of a transcription activator or due to estrangement of transcription suppressors. In this study, it is interesting to find that when the i-motif structure was removed or deformed by either mutation, ASOs, or TMPyP4, the promoter activity was decreased significantly (Figures 3 and 4), suggesting that the i-motif structure in the *BmPOUM2* promoter benefits or enhances its expression.

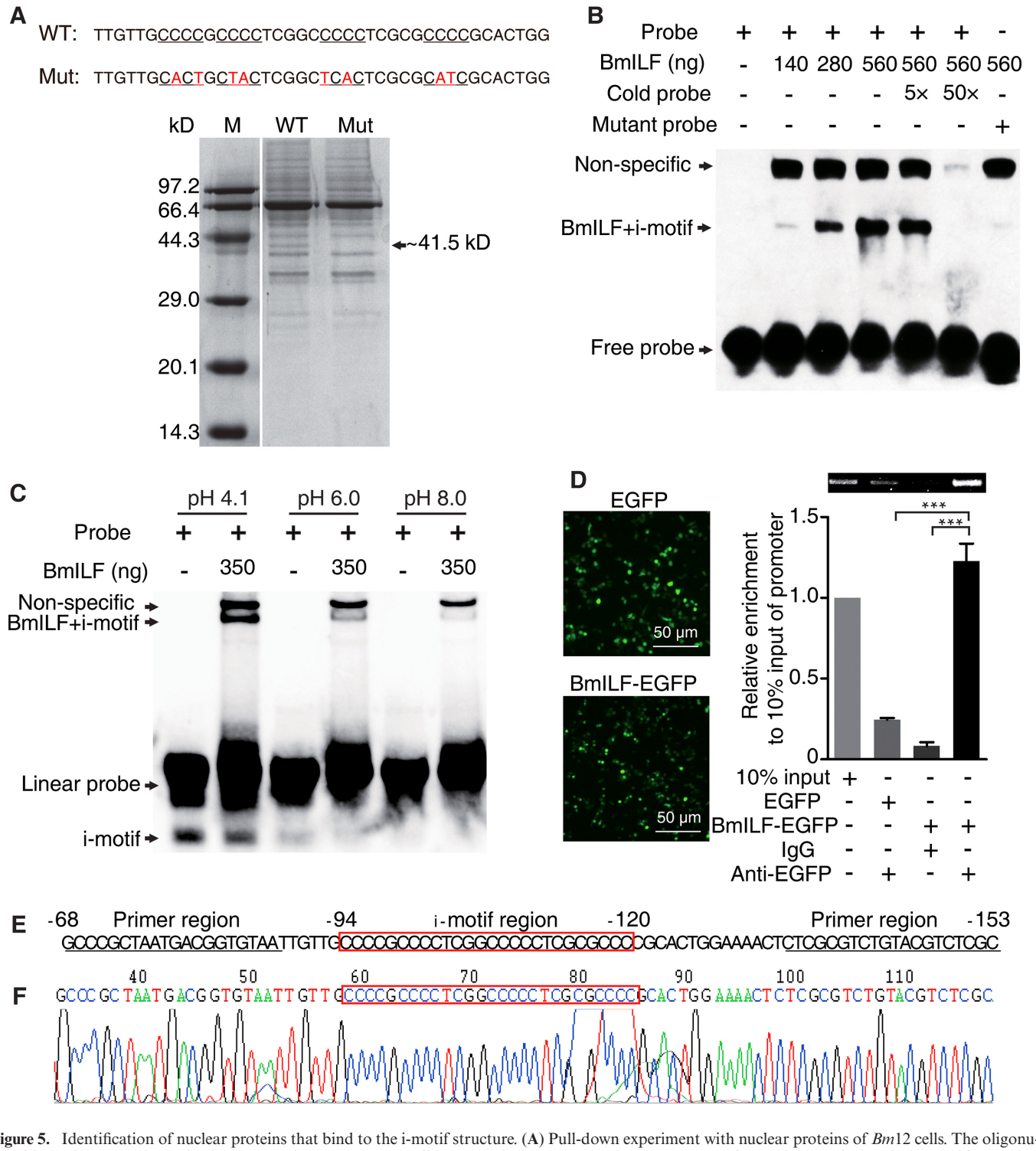

Figure 5. Identification of nuclear proteins that bind to the i-motif structure. (**A**) Pull-down experiment with nuclear proteins of *Bm*12 cells. The oligonucleotide probes were shown in the top panel. WT and Mut: wild-type and mutated ssDNA. The arrows point to the protein band observed in WT but not in the mutant. (**B**) EMSA (6% PAGE) analysis of the recombinant BmILF binding with the i-motif probe. The cold probe was the unlabeled i-motif probe. (**C**) EMSA (12% PAGE) analysis of the recombinant BmILF binding with the i-motif structure or linear DNA probe at different pH conditions. The positions of the labeled i-motif-containing probe, labeled linear DNA probe and the labeled bound i-motif and BmILF are shown by arrows. (**D**) *Bm*12 cells transfected with EGFP vector as a control or with BmILF-EGFP. The green fluorescence shows the similar transfection efficiency and expression of EGFP and BmILF-EGFP. Chromatin immunoprecipitated (ChIP) target sequence was detected by RT-PCR (top panel) and by qRT-PCR (bottom panel). The enrichment of the promoter sequence in immunoprecipitated DNA samples was normalized with DNA present in the 10% input material. Data is means ± SEM ($n = 3$). ***$P < 0.001$ (Student's t test). (**E**) The reverse strand sequence of the -68∼-153 nt region of the *BmPOUM2* promoter. The region that forms the i-motif structure is boxed in red. The primer aligned regions are underlined. (**F**) The sequencing atlas of the enriched RT-PCR product of the ChIP assay.

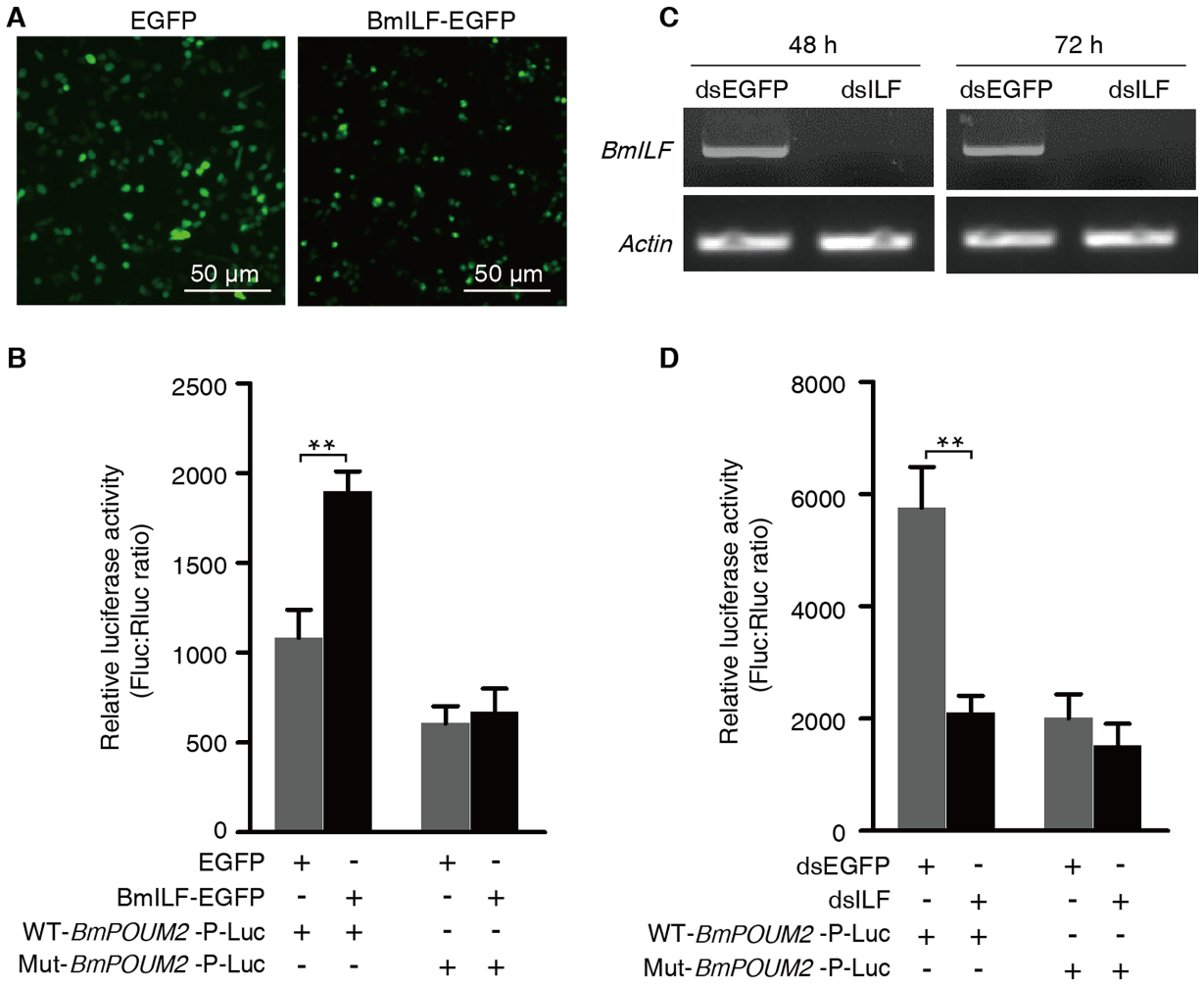

Figure 6. Regulation of the *BmPOUM2* promoter activity by over-expression or RNAi suppression of BmILF. (**A**) *Bm*12 cells co-transfected with EGFP vector as a control or with BmILF-EGFP. The green fluorescence shows the similar transfection efficiency and expression of EGFP and BmILF-EGFP. (**B**) The luciferase activity in *Bm*12 cells co-transfected with EGFP vector or BmILF-EGFP and the pGL3-WT-*BmPOUM2* promoter-luciferase vector (WT-BmPOUM2-P-Luc) or the i-motif-mutated pGL3-Mut-*BmPOUM2* promoter-luciferase vector (Mut-BmPOUM2-P-Luc). (**C**) Reduction of BmILF by RNAi detected at 48 h and 72 h after BmILF dsRNA was transfected into *Bm*12 cells. (**D**) Changes of the luciferase activity in *Bm*12 cells co-transfected with BmILF dsRNA or EGFP dsRNA and the luciferase reporter vector. Data in (B) and (D) are means ± SEM ($n = 3$). $^{**}P < 0.01$ (Student's t test).

The most important finding of this study is that a novel i-motif bound protein, BmILF, was identified and demonstrated by using the methods of DNA-protein pull-down, EMSA and ChIP assays. Although many G4 binding proteins have been reported (60), only one i-motif binding protein (isoform 1 of heterogeneous nuclear ribonucleoprotein L-like, hnRNP LL) was reported so far in detail (31,61,62). hnRNP LL contains four RRM (RNA recognition motif) domains that were predicted to bind with i-motif, resulting in activation of the *bcl-2* gene expression (31,61). BmILF does not have any RRM domain. In human, ILFs have three isoforms, ILF-1 (alias FOXK2), ILF-2 (alias NF45) and ILF-3 (alias NF90), and are involved in regulation of different processes of development and metabolism. The ILF-1 belongs to forkhead gene family (63). The ILF2 and ILF3 both contain a domain associated with zinc fingers (DZF). BmILF also contains a DZF domain and shares about 53% amino acid identity with human ILF2. The human ILF2 and ILF3 can dimerize and act as coactivators to initiate the expression of the *c-fos* gene by association indirectly with RNAPII (RNA polymerase II) (64). However, no study on ILF binding to the i-motif or G4 structure has been reported so far. BmILF is the first identified novel protein that contains a DZF domain and binds with the i-motif structure to activate the expression of *BmPOUM2*.

A regulatory model for the interaction between the i-motif structure and BmILF is suggested based on the data presented in this study (Figure 7). When the transcription of BmPOUM2 is initiated, the duplex DNA strands separate and basal transcriptional factors assemble. The single strand DNA in the GC-rich region forms G4 or i-motif structure due to negative supercoiling. The formed i-motif structure then recruit its binding proteins required for the transcription to activate the transcription complex, turning on the transcription of *BmPOUM2*. After transcription, the basal transcription factor(s) disassemble from the core promoter and the secondary structures deformed to be dsDNA, turning off the transcription.

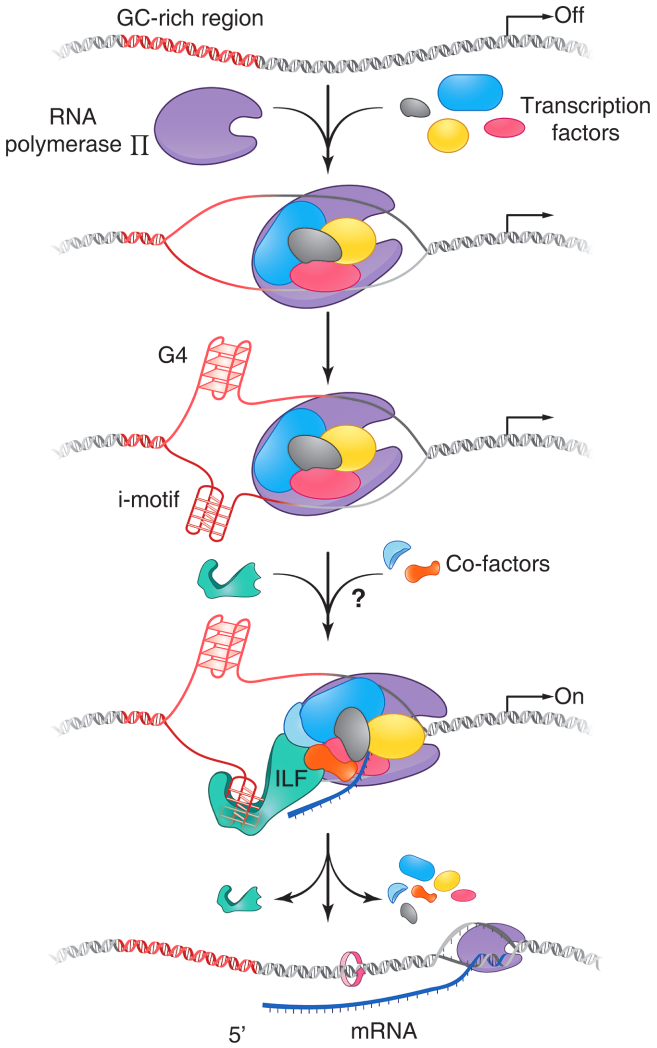

Figure 7. Diagram of proposed epigenetic regulatory mechanism of *BmPOUM2* by the i-motif secondary structure, which was bound by BmILF. When *BmPOUM2* starts to transcribe, the basal transcription factors bind to the core promoter, separating the dsDNA strands. At this time, the GC-rich dsDNA in the upstream region would open because of negative supercoil. The G-rich and C-rich ssDNA immediately fold to form G4 and i-motif structures, respectively. The i-motif recruits the transcription factor BmILF, which may recruit other co-factors required for expression of *BmPOUM2*. After transcription, the basal transcription factor(s) disassemble from the core promoter and the secondary structures unfold to form ds-DNA, turning the transcription off.

One important issue needs to be addressed. Can the i-motif structure actually form and function in *in vivo* physiological condition? This study demonstrated that the i-motif of *BmPOUM2* could *ex vivo* form and bind with BmILF in the cells cultured *in vitro*, regulating the expression of *BmPOUM2*. DNA exists and gene transcription takes place in the nuclei of cells, which have an acidic environment in favor for the formation of i-motif structure. However, further investigation still needs to directly demonstrate the existence and regulatory function of the i-motif structure *in vivo*.

Regulation of gene (or protein) expression can take place at three levels: transcription, post-transcription and translation. The regulation can be genetic (such as gene and promoter primary sequences), epigenetic (for example, methylation and histone modification) or physiological (e.g. mRNA splicing, interference, ribosome assembly). All of these mechanisms involve in only the linear primary sequences with strict A–T and C–G base-pairings. The presence and function of the secondary structures in the DNA and RNA molecules provide a novel epigenetic mechanism for gene transcription regulation. Like methylation and histone modification, formation and deformation of these structures are not directly determined by genetic codes. This mechanism has several typical features. Firstly, it is epigenetic and does not depend on genetic codes of primary sequences. Secondly, the secondary structures stay in a dynamic balance between formation and deformation depending on the status of the cells, in other words, base-pairings can be changed between C–G and G–G/C–C+, which is reversible. Thirdly, the secondary structures can form within ssDNA or mRNA molecules and between two strands of two molecules of DNA or RNA (18). Fourthly, the formation and deformation of the secondary structures may need the involvement of proteins. Once either the formation or deformation of the secondary structures is disturbed by inhibiting the binding proteins, the gene transcription would be suppressed or activated (29,65). This provides a novel target and approach to controlling gene expression, cell function and new pesticides for pest control, gene therapy for diseases, independent of transgenic manipulation.

In summary, this study for the first time identified i-motif structure present in the promoter region of a lepidopteran transcription factor gene in *B. mori*. The i-motif structure, bound with the nuclear transcription factor BmILF, was involved in the transcription regulation of the gene. All these results together provide new insights into the molecular mechanism of the i-motif structure in epigenetic regulation of gene transcription. The mechanism of the *in vivo* formation and function of the i-motif structure in physiological conditions still needs to be further investigated, although some studies have provided some possible mechanisms (26,27).

AVAILABILITY

The amino acid sequence of the transcription factor BmILF is deposited in GenBank (GenBank accession no.: KY082711).

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We would like to thank The Research and Development Center, Sericulture Research Institute, Academy of Agricultural Sciences of Guangdong Province, China, for providing the silkworms and thank Professor Weihua Xu at Sun-Yat Sen University, Guangzhou, China, for helpful discussions and comments on the article.

FUNDING

Chinese National Natural Science Foundation [31330071, 31672494, 31720103916, 2012CB114602]. Funding for open access charge: South China Normal University.
Conflict of interest statement. None declared.

REFERENCES

1. Ruvkun,G. and Finney,M. (1991) Regulation of transcription and cell identity by POU domain proteins. *Cell*, **64**, 475–478.
2. Herr,W., Sturm,R.A., Clerc,R.G., Corcoran,L.M., Baltimore,D., Sharp,P.A., Ingraham,H.A., Rosenfeld,M.G., Finney,M., Ruvkun,G. *et al.* (1988) The POU domain: a large conserved region in the mammalian pit-1, oct-1, oct-2, and *Caenorhabditis elegans* unc-86 gene products. *Genes Dev.*, **2**, 1513–1516.
3. Ryan,A.K. and Rosenfeld,M.G. (1997) POU domain family values: flexibility, partnerships, and developmental codes. *Genes Dev.*, **11**, 1207–1225.
4. Andersen,B. and Rosenfeld,M.G. (2001) POU domain factors in the neuroendocrine system: lessons from developmental biology provide insights into human disease. *Endocr. Rev.*, **22**, 2–35.
5. Komiyama,T., Johnson,W.A., Luo,L. and Jefferis,G.S. (2003) From lineage to wiring specificity POU domain transcription factors control precise connections of *Drosophila* olfactory projection neurons. *Cell*, **112**, 157–167.
6. Certel,K., Hudson,A., Carroll,S.B. and Johnson,W.A. (2000) Restricted patterning of vestigial expression in *Drosophila* wing imaginal discs requires synergistic activation by both Mad and the drifter POU domain transcription factor. *Development*, **127**, 3173–3183.
7. Fukuta,M., Matsuno,K., Hui,C.C., Nagata,T., Takiya,S., Xu,P.X., Ueno,K. and Suzuki,Y. (1993) Molecular cloning of a POU domain-containing factor involved in the regulation of the *Bombyx* sericin-1 gene. *J. Biol. Chem.*, **268**, 19471–19475.
8. Takiya,S., Kokubo,H. and Suzuki,Y. (1997) Transcriptional regulatory elements in the upstream and intron of the fibroin gene bind three specific factors POU-M1, Bm Fkh and FMBP-1. *Biochem, J.*, **321**, 645–653.
9. Johnson,W.A. and Hirsh,J. (1990) Binding of a *Drosophila* POU-domain protein to a sequence element regulating gene expression in specific dopaminergic neurons. *Nature*, **343**, 467–470.
10. Zhang,T.Y., Kang,L., Zhang,Z.F. and Xu,W.H. (2004) Identification of a POU factor involved in regulating the neuron-specific expression of the gene encoding diapause hormone and pheromone biosynthesis-activating neuropeptide in *Bombyx mori*. *Biochem. J.*, **380**, 255–263.
11. Deng,H.M., Zheng,S.C., Yang,X.H., Liu,L. and Feng,Q.L. (2011) Transcription factors BmPOUM2 and BmbFTZ-F1 are involved in regulation of the expression of the wing cuticle protein gene BmWCP4 in the silkworm *Bombyx mori*, *Insect Mol. Biol.*, **20**, 45–60.
12. Deng,H.M., Zhang,J.L., Li,Y., Zheng,S.C., Liu,L., Huang,L.H., Xu,W.H., Palli,S.R. and Feng,Q.L. (2012) Homeodomain transcription factors POU and Abd-A regulate metamorphosis in the silkworm *Bombyx mori*, *Proc. Natl. Acad. Sci. U.S.A.*, **109**, 12598–12603.
13. 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.
14. Sundquist,W.I. and Klug,A. (1989) Telomeric DNA dimerizes by formation of guanine tetrads between hairpin loops. *Nature*, **342**, 825–829.
15. Williamson,J.R., Raghuraman,M.K. and Cech,T.R. (1989) Monovalent cation-induced structure of telomeric DNA: the G-quartet model. *Cell*, **59**, 871–880.
16. Dingley,A.J., Peterson,R.D., Grzesiek,S. and Feigon,J. (2005) Characterization of the cation and temperature dependence of DNA quadruplex hydrogen bond properties using high-resolution NMR. *J. Am. Chem. Soc.*, **127**, 14466–14472.
17. Lipps,H.J. and Rhodes,D. (2009) G-quadruplex structures: in vivo evidence and function. *Trends Cell Biol.*, **19**, 414–422.
18. Rhodes,D. and Lipps,H.J. (2015) G-quadruplexes and their regulatory roles in biology. *Nucleic Acids Res.*, **43**, 8627–8637.
19. Verma,A., Halder,K., Halder,R., Yadav,V.K., Rawal,P., Thakur,R.K., Mohd,F., Sharma,A. and Chowdhury,S. (2008) Genome-wide computational and expression analyses reveal G-quadruplex DNA motifs as conserved cis-regulatory elements in human and related species. *J. Med. Chem.*, **51**, 5641–5649.
20. Huppert,J.L. and Balasubramanian,S. (2007) G-quadruplexes in promoters throughout the human genome. *Nucleic Acids Res.*, **35**, 406–413.
21. Wang,J.M., Huang,F.C., Kuo,M.H., Wang,Z.F., Tseng,T.Y., Chang,L.C., Yen,S.J., Chang,T.C. and Lin,J.J. (2014) Inhibition of cancer cell migration and invasion through suppressing the Wnt1-mediating signal pathway by G-quadruplex structure stabilizers. *J. Biol. Chem.*, **289**, 14612–14623.
22. Seenisamy,J., Rezler,E.M., Powell,T.J., Tye,D., Gokhale,V., Joshi,C.S., Siddiqui-Jain,A. and Hurley,L.H. (2004) The dynamic character of the G-quadruplex element in the c-MYC promoter and modification by TMPyP4. *J. Am. Chem. Soc.*, **126**, 8702–8709.
23. 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. US.A.*, **99**, 11593–11598.
24. David,A.P., Margarit,E., Domizi,P., Banchio,C., Armas,P. and Calcaterra,N.B. (2016) G-quadruplexes as novel cis-elements controlling transcription during embryonic development. *Nucleic Acids Res.*, **44**, 4163–4173.
25. Gehring,K., Leroy,J.L. and Guéron,M.A. (1993) A tetrameric DNA structure with protonated cytosine·cytosine base pairs. *Nature*, **363**, 561–565.
26. 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 (Camb.)*, **46**, 1299–1301.
27. Takahashi,S. and Sugimoto,N. (2015) Pressure-dependent formation of i-motif and G-quadruplex DNA structures. *Phys. Chem. Chem. Phys.*, **17**, 31004–31010.
28. Kendrick,S., Kang,H.J., Alam,M.P., Madathil,M.M., Agrawal,P., Gokhale,V., Yang,D.Z., Hecht,S.M. and Hurley,L.H. (2014) The dynamic character of the BCL2 promoter i-Motif provides a mechanism for modulation of gene expression by compounds that bind selectively to the alternative DNA hairpin structure. *J. Am. Chem. Soc.*, **136**, 4161–4171.
29. Banerjee,K., Wang,M., Cai1,E., Fujiwara1,N., Baker,H. and Cave,J.W. (2015) Regulation of tyrosine hydroxylase transcription by hnRNP K and DNA secondary structure. *Nat. Commun.*, **5**, 5769.
30. 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.
31. 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.
32. Khurad,A.M., Zhang,M.J., Deshmukh,C.G., Bahekar,R.S., Tiple,A.D. and Zhang,C.X. (2009) A new continuous cell line from larval ovaries of silkworm *Bombyx mori*, *In Vitro Cell. Dev. Biol. Anim.*, **45**, 414–419.
33. Sambrook,J., Fritsch,E.F. and Maniatis,T. (1989) *Molecular Cloning: A Laboratory Manual*, 2nd edn. Cold Spring Harbor Laboratory Press, NY.
34. Livak,K.J. and Schmittgen,T.D. (2001) Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) method. *Methods*, **25**, 402–408.
35. Huppert,J.L. and Balasubramanian,S. (2005) Prevalence of quadruplexes in the human genome. *Nucleic Acids Res.*, **33**, 2908–2916.
36. Weitzmann,M.N., Woodford,K.J. and Usdin,K. (1996) The development and use of a DNA polymerase arrest assay for the evaluation of parameters affecting intrastrand tetraplex formation. *J. Biol. Chem.*, **271**, 20958–20964.
37. Han,H., Hurley,L.H. and Salazar,M. (1999) A DNA polymerase stop assay for G-quadruplex-interactive compounds. *Nucleic Acids Res.*, **27**, 537–542.

38. Burge,S., Parkinson,G.N., Hazel,P., Todd,A.K. and Neidle,S. (2006) Quadruplex DNA: sequence, topology and structure. *Nucleic Acids Res.*, **34**, 5402–5415.
39. Cogoi,S. and Xodo,L.E. (2006) G-quadruplex formation within the promoter of the KRAS proto-oncogene and its effect on transcription. *Nucleic Acids Res.*, **34**, 2536–2549.
40. Kikin,O., D'Antonio,L. and Bagga,P.S. (2006) QGRS Mapper: a web-based server for predicting G-quadruplexes in nucleotide sequences. *Nucleic Acids Res.*, **34**, W676–W682.
41. Ambrus,A., Chen,D., Dai,J., Jones,R.A. and Yang,D. (2005) Solution structure of the biologically relevant G-quadruplex element in the human c-myc promoter. Implications for G-quadruplex stabilization. *Biochemistry*, **44**, 2048–2058.
42. Xu,Y. and Sugiyama,H. (2006) Formation of the G-quadruplex and i-motif structures in retinoblastoma susceptibility genes (Rb). *Nucleic Acids Res.*, **34**, 949–954.
43. Sun,D., Guo,K., Rusche,J.J. and Hurley,L.H. (2005) Facilitation of a structural transition in the polypurine/polypyrimidine tract within the proximal promoter region of the human VEGF gene by the presence of potassium and G-quadruplex-interactive agents. *Nucleic Acids Res.*, **33**, 6070–6080.
44. Rankin,S., Reszka,A.P., Huppert,J., Zloh,M., Parkinson,G.N., Todd,A.K., Ladame,S., Balasubramanian,S. and Neidle,S. (2005) Putative DNA quadruplex formation within the human c-kit oncogene. *J. Amer. Chem. Soc.*, **127**, 10584–10589.
45. Fernando,H., Reszka,A.P., Huppert,J., Ladame,S., Rankin,S., Venkitaraman,A.R., Neidle,S. and Balasubramanian,S. (2006) A conserved quadruplex motif located in a transcription activation site of the human c-kit oncogene. *Biochemistry*, **45**, 7854–7860.
46. De Armond,R., Wood,S., Sun,D., Hurley,L.H. and Ebbinghaus,S.W. (2005) Evidence for the presence of a guanine quadruplex forming region within a polypurine tract of the hypoxia inducible factor 1alpha promoter. *Biochemistry*, **44**, 16341–16350.
47. Dexheimer,T.S., Sun,D. and Hurley,L.H. (2006) Deconvoluting the structural and drug-recognition complexity of the G-quadruplex-forming region upstream of the bcl-2 P1 promoter. *J. Am. Chem. Soc.*, **128**, 5404–5415.
48. Qin,Y., Rezler,E.M., Gokhale,V., Sun,D. and Hurley,L.H. (2007) Characterization of the G-quadruplexes in the duplex nuclease hypersensitive element of the PDGF-A promoter and modulation of PDGF-A promoter activity by TMPyP4. *Nucleic Acids Res.*, **35**, 7698–7713.
49. Guo,K., Gokhale,V., Hurley,L.H. and Sun,D. (2008) Intramolecularly folded G-quadruplex and i-motif structures in the proximal promoter of the vascular endothelial growth factor gene. *Nucleic Acids Res.*, **36**, 4598–4608.
50. Dettler,J.M., Buscaglia,R., Cui,J., Cashman,D., Blynn,M. and Lewis,E.A. (2010) Biophysical characterization of an ensemble of intramolecular i-motifs formed by the human c-MYC NHE III1 P1 promoter mutant sequence. *Biophys. J.*, **99**, 561–567.
51. Kumar,N., Patowary,A., Sivasubbu,S., Petersen,M. and Maiti,S. (2008) Silencing c- MYC expression by targeting quadruplex in P1 promoter using locked nucleic acid trap. *Biochemistry*, **47**, 13179–13188.
52. Fedoroff,O.Y., Rangan,A., Chemeris,V.V. and Hurley,L.H. (2000) Cationic porphyrins promote the formation of i-motif DNA and bind peripherally by a nonintercalative mechanism. *Biochemistry*, **39**, 15083–15090.
53. Khan,N., Aviñó,A., Tauler,R., González,C., Eritja,R. and Gargallo,R. (2007) Solution equilibria of the i-motif-forming region upstream of the B-cell lymphoma-2 P1 promoter. *Biochimie*, **89**, 1562–1572.
54. Grand,C.L., Han,H., Munoz,R.M., Weitman,S., Von Hoff,D.D., Hurley,L.H. and Bearss,D.J. (2002) The cationic porphyrin TMPyP4 down-regulates c-MYC and human telomerase reverse transcriptase expression and inhibits tumor growth *in vivo*. *Mol. Cancer Ther.*, **1**, 565–573.
55. Suetsugu,Y., Futahashi,R., Kanamori,H., Kadono-Okuda,K., Sasanuma,S., Narukawa,J., Ajimura,M., Jouraku,A., Namiki,N., Shimomura,M. *et al.* (2013) Large scale full-length cDNA sequencing reveals a unique genomic landscape in a Lepidopteran model insect *Bombyx mori*. *G3 (Bethesda)*, **3**, 1481–1492.
56. de Folter,S., Urbanus,S.L., van Zuijlen,L.G., Kaufmann,K. and Angenent,G.C. (2007) Tagging of MADS domain proteins for chromatin immunoprecipitation. *BMC Plant Biol.*, **7**, 47.
57. Chambers,V.S., Marsico,G., Boutell,J.M., Di Antonio,M., Smith,G.P. and Balasubramanian,S. (2015) High-throughput sequencing of DNA G-quadruplex structures in the human genome. *Nat. Biotechnol.*, **33**, 877–881.
58. Hänsel-Hertsch,R., Beraldi,D., Lensing,S.V., Marsico,G., Zyner,K., Parry,A., Di Antonio,M., Pike,J., Kimura,H., Narita,M. *et al.* (2016) G-quadruplex structures mark human regulatory chromatin. *Nat. Genet.*, **48**, 1267–1272.
59. Shklover,J., Weisman-Shomer,P., Yafe,A. and Fry,M. (2010) Quadruplex structures of muscle gene promoter sequences enhance in vivo MyoD-dependent gene expression. *Nucleic Acids Res.*, **38**, 2369–2377.
60. Fry,M. (2007) Tetraplex DNA and its interacting proteins. *Front Biosci.*, **12**, 4336–4351.
61. Roy,B., Talukder,P., Kang,H.J., Tsuen,S.S., Alam,M.P., Hurley,L.H. and Hecht,S.M. (2016) Interaction of individual structural domains of hnRNP LL with the BCL2 promoter i-motif DNA. *J. Am. Chem. Soc.*, **138**, 10950–10962.
62. Day,H.A., Pavlou,P. and Waller,Z.A. (2014) i-Motif DNA: structure, stability and targeting with ligands. *Bioorg. Med. Chem.*, **22**, 4407–4418.
63. Fujii,Y. and Nakamura,M. (2010) FOXK2 transcription factor is a novel G/T-mismatch DNA binding protein. *J. Biochem.*, **147**, 705–709.
64. Nakadai,T., Fukuda,A., Shimada,M., Nishimura,K. and Hisatake,K. (2015) The RNA binding complexes NF45-NF90 and NF45-NF110 associate dynamically with the c-fos gene and function as transcriptional coactivators. *J. Biol. Chem.*, **290**, 26832–26845.
65. Qiu,J., Chen,S.Q., Su,L.J., Liu,J.G., Xiao,N.N., Ou,T.M., Tan,J.H., Gu,L.Q., Huang,Z.S. and Li,D. (2014) Cellular nucleic acid binding protein suppresses tumor cell metastasis and induces tumor cell death by downregulating heterogeneous ribonucleoprotein K in fibrosarcoma cells. *Biochim. Biophys. Acta*, **1840**, 2244–2252.