

Non-canonical Structures in Promoter Modulate Gene Expression in *Escherichia coli*

Sabina Božič Abram,¹ Maja Marušič,¹ Martina Lenarčič Živković,¹ Jasna Brčić,¹ Janez Plavec^{1,2,3,*}

¹ Slovenian NMR Centre, National Institute of Chemistry, Hajdrihova 19, SI-1000 Ljubljana, Slovenia

² EN-FIST Centre of Excellence, Trg OF 13, SI-1000 Ljubljana, Slovenia

³ Faculty of Chemistry and Chemical Technology, University of Ljubljana, Večna pot 113, SI-1000 Ljubljana, Slovenia

* Corresponding author's e-mail address: janez.plavec@ki.si

RECEIVED: April 18, 2018 * REVISED: May 9, 2018 * ACCEPTED: May 10, 2018

THIS PAPER IS DEDICATED TO DR. BISERKA KOJIĆ-PRODIĆ ON THE OCCASION OF HER 80TH BIRTHDAY

Abstract: Herein we show how sequences that can form different non-canonical structures affect gene expression levels when inserted in the core of σ^{70} -dependent promoter, between the –35 and –10 elements recognized by RNA polymerase, in *E. coli*. We note that influence on level of GFP expression varies considerably depending on introduction of non-canonical structural elements in the antisense and sense strands as well as with their propensities to form G-triplex, G-hairpin, hairpin or G-quadruplex structures. Moreover, the extent of repression of expression does not relate to the *in vitro* thermal stability in a simple manner. Repression is most likely caused by steric interference rather than improper distance between the –35 and –10 elements. Although properties like thermal stability and topology can be somewhat different under *in vivo* and *in vitro* conditions, our results suggest that the extent of expression suppression cannot be dependent solely on thermal stabilities of G-rich structures alone.

Keywords: DNA, non-canonical structures, topology, gene expression modulation, *E. coli* σ^{70} , GFP expression.

INTRODUCTION

INITIATION of gene transcription is a tightly controlled process that begins with binding of RNA polymerase to specific sequences in the promoter region of DNA. Recognition and binding to the promoter sequence depends on the σ subunit of the RNA polymerase holoenzyme.^[1] The primary σ factor of *E. coli* is σ^{70} , which is essential for expression of genes required for normal cell growth. The σ^{70} subunit of RNA polymerase in a sequence-specific manner interacts with consensus hexamer DNA sequences of the –35 and –10 promoter elements, which are located upstream from the transcription start site (TSS). The highest gene expression levels are achieved when the –35 and –10 elements are separated by 17 nucleotides (nt) in double-stranded DNA form. While RNA polymerase can only tolerate about ± 1 nt difference in length between the –35 and –10 elements, its activity is not sensitive with respect to the sequence of the region in between.^[1–4]

G-rich DNA sequences can, in addition to the canonical double-stranded structure, adopt a great variety of non-canonical structures. Some of the most studied non-canonical structures are G-quadruplexes (GQ), which have been implicated in important biological processes. Potential GQ-forming sequences, which are characterized by short G-rich tracts connected by regions that vary in length and sequence, are found at telomeric repeats, promoter regions of many genes and human proto-oncogenes.^[5–9] While it has been established that GQ formation can affect gene expression, the precise mechanism remains elusive.^[10–14] Moreover, in recent study by Takahashi *et al.* it was shown that biological functions of GQ in the complex intracellular environment can also be regulated by oxidative lesions that change GQ structure and thus decrease its stability.^[15] Four-stranded GQ structures are composed of stacked layers of G-quartets, formed by four guanine residues connected by Hoogsteen-type hydrogen bonds and stabilized by monovalent cations. GQs can adopt parallel, antiparallel

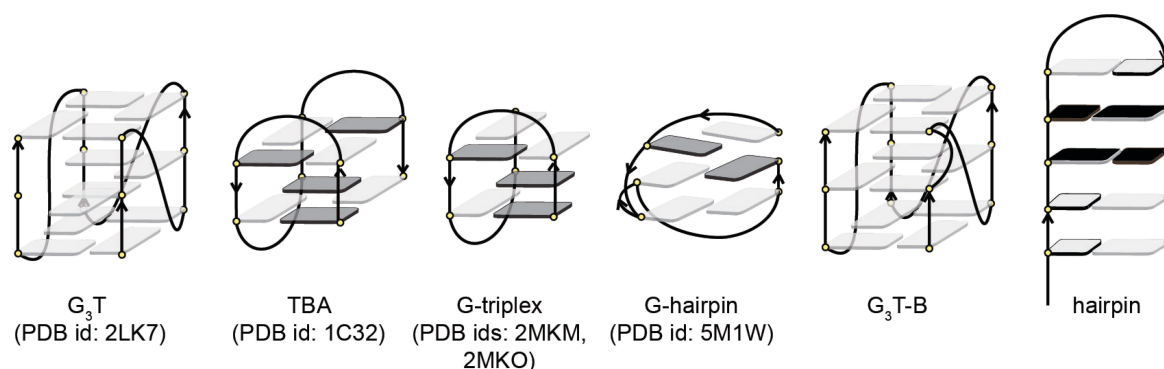


Figure 1. Schematic representation of non-canonical structures used in this study with respective high-resolution structures: parallel GQ G_3T ,^[21] antiparallel GQ TBA,^[22] G-triplex,^[18,19] G-hairpin,^[17] parallel GQ with a bulge (G_3T-B)^[23] and hairpin. Guanines in *anti* and *syn* conformation around the glycosidic bond are designated as light and dark grey rectangles, respectively. Adenines are designated as black rectangles, while cytosines and thymines are labelled with light grey and black squares, respectively.

and hybrid (3+1) topologies characterized by different orientation of the four strands.^[16] In addition to GQs, G-rich sequences can also adopt other non-canonical structures, such as a G-hairpin, a fold-back structure stabilized by G:G base pairs^[17] and a G-triplex, three stranded structure stabilized by G:G:G triad planes (Figure 1).^[18,19]

It has been established that processivity of DNA polymerase is affected by stability and topology of non-canonical structures formed during replication.^[20] Similarly, non-canonical structures in the promoter region were shown to affect the processivity of RNA polymerase.^[10,12] We investigated *if* and *how* different sequences with the ability to form non-canonical structures affect gene expression when positioned in the region between -35 and -10 elements of the constitutive σ^{70} – dependent promoter of a reporter gene coding for GFP (green fluorescent protein) in *E. coli*. Formation of a non-canonical structure in the region between -35 and -10 promoter elements was expected to affect gene expression only at the level of transcription. Sequences that adopt diverse topologies (Figure 1) with different thermal stabilities *in vitro* were selected. Structure formation can affect gene expression in two different ways. It could either represent a steric barrier which interferes with binding of RNA polymerase or disrupt the optimal length between the -35 and -10 promoter elements that represent RNA polymerase binding sites. To discern if the effect of non-canonical structure formation between -35 and -10 elements on gene expression is due to steric effects or the difference in length between, we systematically added nucleotides up- and downstream of the GQ-forming sequence. It was expected that compensating for the length difference between the two elements would restore RNA polymerase activity.

EXPERIMENTAL SECTION

Materials, Plasmids and *E. coli* Strain

Enzymes were purchased from New England Biolabs. Oligonucleotides were synthesized by IDT (USA) at the 4 nM scale. Experiments were conducted with the *E. coli* Rosetta (DE3) strain (Novagen; F- *ompT hsdSB(rB- mB-) gal dcm* (DE3) pRARE (CamR)). Plasmids were introduced by heat-shock transformation. The promoterless plasmid pAcGFP1-1 was purchased from Takara-Clontech (USA). The constitutive promoter was from the J06 promoter from the Anderson promoter library used in Holder *et al.*^[10] (see Ref. 10 for appropriate link) Luria – Bertani (LB) medium and agar were purchased from Sigma – Aldrich.

Molecular Cloning Procedure

Standard cloning procedures were performed. Briefly, oligonucleotides (IDT, USA) with EcoRI and AgeI cohesive ends were used as inserts for the EcoRI/ AgeI digested pAcGFP1-1 plasmid. The oligonucleotides were first hybridized by temperature annealing from 95 °C to 25 °C within 120 min. Afterwards they were ligated into purified EcoRI/ AgeI digested plasmid pAcGFP1-1 with the Quick ligation kit (NEB) according to the manufacturer's instructions. Sequences of all constructs were verified by Sanger sequencing at GATC Biotech.

GFP Expression Assay

Assembled constructs were transformed into chemically competent Rosetta (DE3) *E. coli* strain and plated onto LB plates supplemented with 50 µg/ml kanamycin and 30 µg/l chloramphenicol. Colonies carrying constructs with non-canonical DNA structures in the promoter region and GFP protein as a gene expression reporter were inoculated into

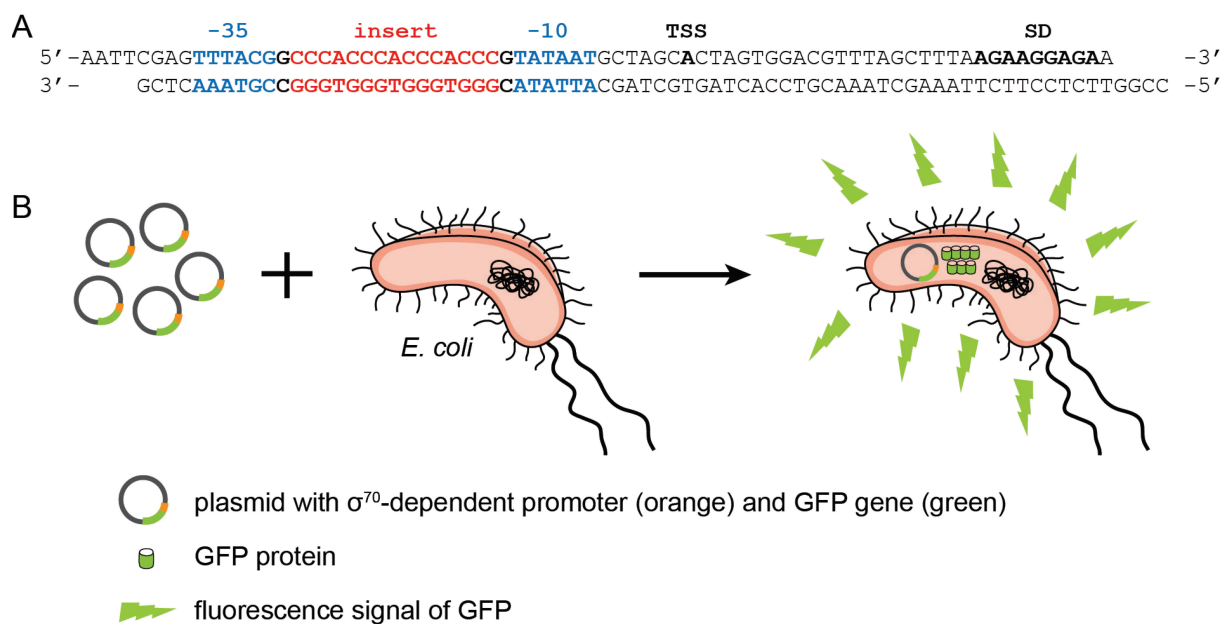


Figure 2. Schematic presentation of experimental design. A) Sequences expected to form different non-canonical structures (red) were inserted between -35 and -10 elements (blue) of the constitutive σ^{70} -dependent promoter with GFP expression system. An example where the structure-forming sequence was introduced in the antisense strand is shown here. B) Plasmid with the constitutive σ^{70} -dependent promoter and GFP gene were transformed into *E. coli* expression strain. GFP fluorescence levels were measured after overnight incubation and compared to respective controls.

liquid LB medium supplemented with 50 $\mu\text{g/ml}$ kanamycin and 30 $\mu\text{g/l}$ chloramphenicol and grown at 37 $^{\circ}\text{C}$ and 180 rpm. 1 ml of cell suspension was then pelleted by centrifugation at 13 000 rpm and resuspended in PBS buffer. Two hundred microliters of each cell culture suspension were then transferred into a 96-well plate and GFP fluorescence and optical density were analyzed on a microplate reader Synergy Mx (BioTek; excitation wavelength = 475 nm, emission wavelength = 505 nm). Fluorescence values were divided with the corresponding optical density measured at 600 nm. All experiments were conducted in triplicates, vertical bars represent standard deviation.

UV Measurements

UV measurements were performed on a Varian CARY-100 BIO UV-VIS spectrophotometer with the Cary Win UV Thermal program using a 1 cm path-length cells. Cooling and heating curves were obtained at a rate of 0.3 $^{\circ}\text{C}/\text{min}$ from 10 $^{\circ}\text{C}$ to 90 $^{\circ}\text{C}$ with the starting point at 90 $^{\circ}\text{C}$. Measurements were repeated three times and absorbance was measured at 260 and 295 nm at 0.5 $^{\circ}\text{C}$ steps. The concentration of samples was 10 μM per strand. Samples were dissolved in 10 mM potassium phosphate buffer with addition of 70 mM KCl. A combination of mineral oil and a fixed cuvette cap was used to prevent evaporation and

sample loss at high temperatures. A stream of nitrogen was applied to prevent condensation at low temperatures.

RESULTS

Gene Expression is affected by the Sequence inserted into the Core Promoter

We set to examine how six sequences that were shown to form different non-canonical structures affect gene expression when they are inserted in the core promoter, a region of consensus length of 17 nt between the conserved hexameric sequences of -35 and -10 promoter elements (Figure 2A). Sequences were inserted in a plasmid and introduced into an *E. coli* expression strain, with the structure-forming sequence located either in the antisense or sense strand. Transcription efficiency was assessed using GFP expression system (Figure 2B). Expression levels of six sequences that were expected to form various non-canonical structures were compared with respective control sequences.

We inserted sequences that were previously shown to fold *in vitro* into a very stable parallel GQ (G_3T),^[21] an antiparallel GQ (TBA, thrombin-binding aptamer),^[22] a G-triplex,^[18,19] a G-hairpin^[17] and a parallel GQ with a bulge ($\text{G}_3\text{T-B}$)^[23] (Figure 1, Table 1). For comparison, we also

Table 1. Deoxynucleotide sequences and respective T_m values and structures formed *in vitro*

NAME	INSERTED SEQUENCE (5'→3') ^(a)	T_m ^(b) /°C	IN VITRO STRUCTURE
CTRL1	C TTGTTTGGTTTGGTTT C	–	dsDNA
G ₃ T	C GGGTGGGTGGGTGGG C	>95	parallel GQ
TBA	C GGTGGGTGGTGGG C	49	antiparallel GQ
G-triplex	C GATGGTTGGTGGG C	34 ^(c)	G-triplex
G-hairpin	C GATGTGTGGGTGTGC C	53 ^(d)	G-hairpin
hairpin	C TCCATGTTTCATGG C	52	hairpin
CTRL2	C TTTGTTTGGTTTGGTTT C	–	dsDNA
G ₃ T-B	C GGTGTGGGTGGGTGGG C	79	parallel GQ with a bulge

^(a) Inserted sequences with the ability to form a non-canonical structure (and controls) are shown in red whereas spacer cytosines are shown in black.

^(b) T_m values are given for 10 μ M DNA in 10 mM potassium phosphate buffer, pH 7 and 70 mM KCl. T_m values of control sequences were not determined.

^(c) Reported by Limongelli *et al.*^[18] for 0.2 mM DNA in 10 mM potassium phosphate buffer, pH 7, 70 mM KCl and 0.2 mM EDTA.

^(d) Reported by Gajarsky *et al.*^[17] for 0.31 mM DNA in 10 mM potassium phosphate buffer, pH 7, 100 mM KCl and 0.2 mM EDTA.

inserted a sequence which *in vitro* forms a hairpin structure stabilized with Watson-Crick base pairs (Supplementary Figure 1).

The sequences were first introduced into the antisense strand. Inserted sequences consisted of altogether 17 nt which included a 15 nt insert with the ability to form a non-canonical structure (labelled red in Table 1) and one cytosine residue on each side of the 15 nt insert (labelled black in Table 1). Cytosines were introduced to serve as a spacer between the two RNA polymerase binding sites and 15 nt inserts. A 15 nt sequence named CTRL1 was designed as a control and was predicted to form canonical double-stranded DNA structure (dsDNA). Since the sequence G₃T-B that was predicted to fold into a

parallel GQ with a bulge was 1 nt longer, 16 nt long sequence CTRL2 was designed to serve as a control.

It has been shown earlier that effect on gene expression is strongly dependent on whether the GQ-forming sequence is located in the antisense or sense strand.^[10,12,24] Therefore, we tested also the effect of sequences able to form a parallel GQ, an antiparallel GQ and a hairpin structure on expression levels when the sequences are inserted in the sense strand (Figure 3A).

Gene expression levels of GFP decreased by about 50 % when sequences G₃T and TBA, which are able to form a parallel and antiparallel GQ, respectively, were introduced in the antisense stand (Figure 4). Fluorescence levels for the hairpin-forming sequence were 40 % lower in comparison

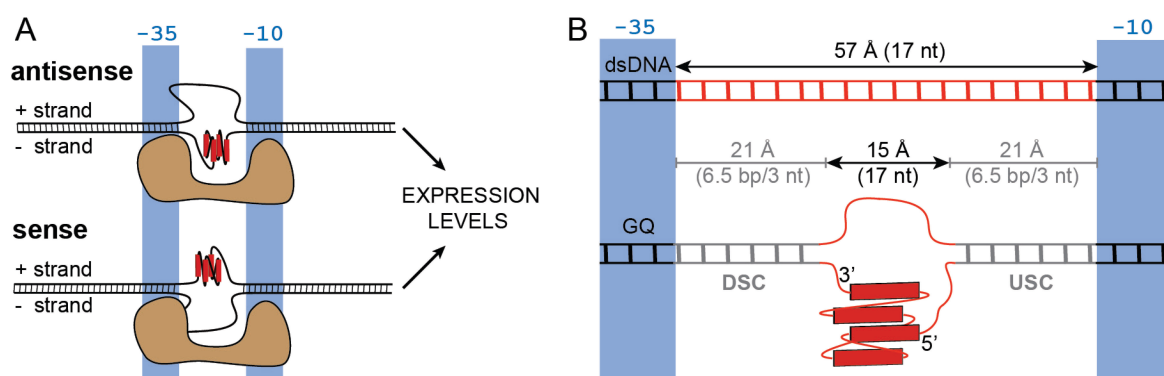


Figure 3. Effect of sequences that can form non-canonical structures on gene expression. A) Sequences able to form different non-canonical structures were introduced in the antisense and the sense strand. Transcription was monitored by GFP expression level. B) Schematic presentation of length compensation experiment. The 17 nt region between the –35 and –10 elements that forms dsDNA or GQ is marked with red. The 5' and 3' ends of the GQ are marked. Nucleotides of the spacer regions are coloured grey. Nucleotides inserted closer to the –35 element and to the 3' end are marked with DSC (downstream compensation), while nucleotides inserted closer to the –10 element and to the 5' end are marked with USC (upstream compensation).

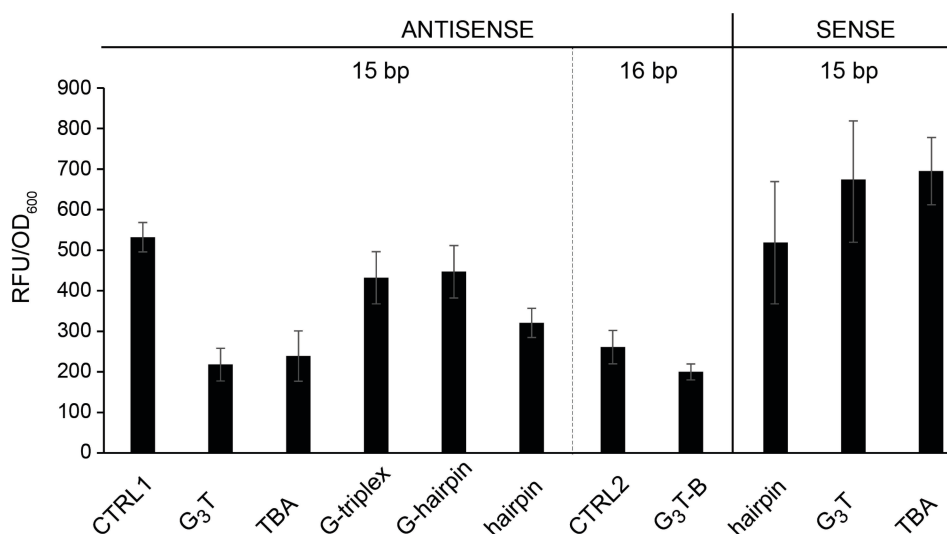


Figure 4. GFP expression levels of sequences able to form different non-canonical structures and control sequences. The fluorescence intensity is represented as average value of three independent measurements with the standard deviation shown as vertical bars.

to CTRL1, whereas minor, if any effect on gene expression was observed for G-triplex and G-hairpin sequences in the antisense strand. Insertion of the G₃T-B sequence resulted in a 30 % decrease of the GFP gene expression level relative to its 16 nt long control sequence CTRL2.

The presence of G₃T and TBA in the sense strand increased gene expression by 20 % (Figure 4). In comparison, insertion of the hairpin-forming sequence in the sense strand did not significantly affect gene expression relative to the control sequence CTRL1.

Length Compensation Experiment

The ideal length of the sequence between the -35 and -10 elements is 17 nt, which corresponds to about 57 Å when folded into dsDNA (Figure 3B). The length can shorten significantly if the sequence between the -35 and -10 elements folds into a non-canonical structure. For example, the distance between 5'- and 3'-ends of the parallel GQ formed by G₃T (PDB id 2LK7) is around 15 Å. Shorter distance between the -35 and -10 elements could by itself be detrimental for RNA polymerase activity. We expected that adding nucleotides (spacers) in the region up- and downstream from the G-rich sequence will compensate for the difference in length due to GQ formation. Nucleotides in the spacer regions up- and downstream from the GQ-forming sequence can form dsDNA or remain unstructured. Reduced distance of around 42 Å (21 Å on each side of the G₃T sequence) can be compensated by overall number of 13 nt, if the nucleotides in the spacer regions form dsDNA (rise per base pair in B-form DNA is cca. 3.3 Å)^[25] or 6 nt, if

the nucleotides in the spacer regions are unstructured (around 7.0 Å per nt, measured in GQ structures with PDB ids 2KPR and 1U64).^[26,27] To account for both possibilities for nucleotides in the spacer region (*i.e.* dsDNA and unstructured), we systematically added between three and eight nucleotides up- and downstream of the G₃T sequence, with overall number of added nucleotides in the range between seven and fifteen (overall number is reflected in the names of the sequences in Table 2).

We observed that adding spacer nucleotides to the G₃T sequence does not restore expression levels relative to CTRL1 sequence (Figure 5). Interestingly, the GFP fluorescence levels of G₃T-extended sequences were about 70 % lower than in the case of G₃T sequence.

DISCUSSION

We examined how sequences with the ability to form distinct non-canonical structures affect gene expression when they are inserted between the -35 and -10 elements of the σ^{70} -dependent promoter of GFP reporter gene in *E. coli*. Decrease in gene expression was observed for parallel GQ-, antiparallel GQ- and hairpin-forming sequences when they are placed in the antisense strand. Sequences able to form G-hairpin and G-triplex structures inserted in the antisense strand did not affect GFP expression levels. It may be expected that G-hairpin structure failed to form in a context of a longer DNA sequence *in vivo*. This is most likely a consequence of its structural features, since G-hairpin is a fold-back structure characterized by stacking of the 5'-

Table 2. Deoxynucleotide sequences designed to compensate for the length difference between the –35 and –10 promoter elements

NAME	INSERTED SEQUENCE (5'→3') ^(a)	LENGTH/nt	USC ^(b) /nt	DSC ^(c) /nt
G ₃ T ₊₇	C TTT GGGTGGGTGGGTGGG TTTT C	22	3	4
G ₃ T ₊₈	C TCTT GGGTGGGTGGGTGGG TTCT C	23	4	4
G ₃ T ₊₉	C CTCTT GGGTGGGTGGGTGGG TTCT C	24	4	5
G ₃ T ₊₁₀	C CTCTT GGGTGGGTGGGTGGG TCTTC C	25	5	5
G ₃ T ₊₁₁	C CTCTT GGGTGGGTGGGTGGG TTCTTC C	26	5	6
G ₃ T ₊₁₂	C CTCTT GGGTGGGTGGGTGGG TTCTTC C	27	6	6
G ₃ T ₊₁₃	C CTCTT GGGTGGGTGGGTGGG TTCTTCC C	28	6	7
G ₃ T ₊₁₄	C CCTCTT GGGTGGGTGGGTGGG TTCTTCC C	29	7	7
G ₃ T ₊₁₅	C CCTCTT GGGTGGGTGGGTGGG TTCTTCTC C	30	7	8

^(a) Inserted sequences with the ability to form a non-canonical structure (and controls) are shown in red whereas spacer cytosines are shown in black and DSC and USC deoxynucleotides are shown in grey.

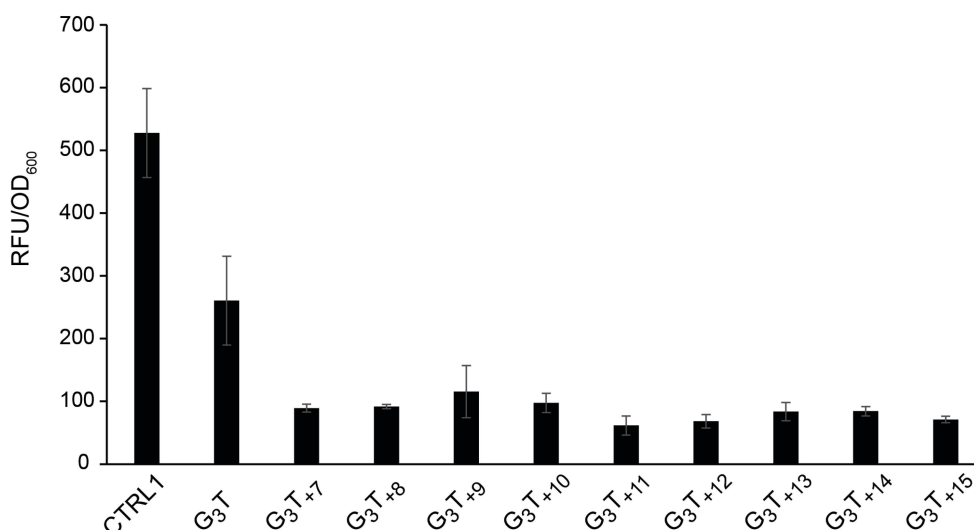
^(b) USC, upstream compensation: number of nucleotides inserted closer to the 5'-end of the antisense strand of DNA.

^(c) DSC, downstream compensation: number of nucleotides inserted closer to the 3'-end of the antisense strand of DNA.

and 3'-terminal residues.^[17] Similarly, G-triplex formation *in vitro* was only demonstrated for a truncated GQ-forming sequence and was disrupted if nucleotides were added at the 3'-end of the sequence.^[18] Our results suggest that sequences inserted in the antisense strand affected gene expression only when corresponding non-canonical structures were able to form in the context of longer DNA sequence in a plasmid (*i.e.* in the case of parallel GQ-, antiparallel GQ- and hairpin-forming sequence). Comparable levels of gene expression were detected although the structures are topologically very different. This suggests that the presence and not topology of non-canonical structure importantly affect the level of expression when the structure-forming sequence is inserted between the –35 and –10 elements of the σ^{70} -

dependent promoter in *E. coli*. Our observations are in agreement with a recent study which showed that the GQ topology does not significantly affect the level of *in vitro* transcription by T7 RNA polymerase when GQ-forming sequences are placed in the T7 promoter.^[13]

Moreover, our results show that gene expression levels cannot be correlated in a simple manner with thermal stabilities of non-canonical structures determined *in vitro*. These findings might supplement results in a recent study by the Sugimoto group which found that transcriptional level strongly depends on thermodynamic stability of the GQ structure when GQ-forming sequence is inserted downstream of the core promoter.^[12] However, this apparent discrepancy highlights the importance that the exact location of GQ-forming sequence within the

**Figure 5.** GFP expression levels of CTRL1, G3T and G3T-extended sequences. The fluorescence intensity is represented as average value of three independent measurements with the standard deviation shown as vertical bars.

promoter plays in determining its effect on gene expression, as was established earlier by the Hartig group.^[10]

Additionally, we observed that gene expression is affected differently when GQ- or hairpin-forming sequences are located in the antisense or in the sense strand. Distinct effect on expression level when placing the GQ-forming sequence on antisense and sense strand has been observed before.^[10,24,28,29] In contrast to the effect of GQ- and hairpin-forming sequences in the antisense strand, they caused a slight increase or no change in expression levels when located in the sense strand. Similarly, a study by the Hartig group showed that parallel GQ in the core promoter region of the sense strand has no effect on gene expression.^[10] A slight increase in expression level was observed for GQ-forming sequences G₃T and TBA when they were inserted in the sense strand. Although the structure in the sense strand does not interfere directly with RNA polymerase binding, formation of a non-canonical structure in the sense strand might simplify separation of the strands of double stranded DNA and thus facilitate expression. On the other hand, as reported recently, a formal sequence reversal of the G-rich DNA segment from 5'-3' to 3'-5' exerts a substantial effect on the number of structures formed, while the type of GQ fold is in fact determined by the presence of K⁺ or Na⁺ ions.^[30] Importantly, the melting temperatures of GQs adopted by oligonucleotides with sequences in the 5'-3' direction are higher than those of their 3'-5' counterparts with both KCl and NaCl.

In addition to representing a steric barrier, GQ formation could hamper the activity of RNA polymerase by disrupting the distance between -35 and -10 elements which are binding sites for RNA polymerase. In order to discern between steric effect and difference in length, we systematically added nucleotides up- and downstream of the G₃T sequence. We expected that such spacers would compensate for the difference in length due to GQ structure formation and regain expression levels. Although spacers with different lengths were added, expression levels were not restored. Moreover, expression levels of sequences with added nucleotides were even lower than of G₃T sequence alone. This indicates that compensating for difference in length between the -35 and -10 elements is not sufficient to restore expression levels. Furthermore, if the repression of the gene expression would be caused by improper distance alone, we would expect that G₃T formation in the sense strand would cause a similar decrease in expression level instead of a slight increase that was observed. Underlying mechanism of gene expression repression is thus most likely caused by steric interference with RNA polymerase recognition and binding.

CONCLUSIONS

Effect of sequence variation of DNA regions that are able to form different non-canonical structures and thus affect expression levels in *E. coli* was studied when placed in the antisense or sense strand of the core promoter region of RNA polymerase. In the antisense strand, the extent of expression level is decreased by the presence of non-canonical structures. However, in contrast to earlier studies we found that *in vitro* thermal stabilities cannot be related to the level of repression of the gene expression in a simple manner. For example, in the case of G₃T and TBA a more efficient suppression of expression would be expected for G₃T due to its substantially higher *in vitro* determined melting temperature in comparison to TBA. However, TBA inhibits expression to similar extent as G₃T suggesting that antiparallel topology of TBA must play an important role in the way repression of the expression occurs, possibly owing to faster folding or slower unfolding in comparison to a parallel stranded G₃T. In the sense strand, a slight increase in expression levels was observed. As formation of non-canonical structures disrupts the length between the -35 and -10 promoter elements recognized by RNA polymerase we introduced different spacers for length compensation, but expression levels could not be restored. Taken together, repression of the gene expression is most likely caused by steric interference, rather than improper distance between the -35 and -10 elements of σ^{70} -dependent promoter in *E. coli*. Although properties like thermal stability and topology can be somewhat different under *in vivo* and *in vitro* conditions, our results suggest that the extent of expression suppression cannot be dependent solely on thermal stabilities of G-rich structures alone.

Acknowledgment. The authors acknowledge the financial support from Slovenian Research Agency (research core funding No. P1-242, J1-6733 and J3-7245).

Supplementary Information. Supporting information to the paper is attached to the electronic version of the article at: <https://doi.org/10.5562/cca3349349>.

REFERENCES

- [1] Y. Zuo, T. A. Steitz, *Mol. Cell.* **2015**, *58*, 534.
- [2] A. J. Dombroski, B. D. Johnson, M. Lonetto, C. A. Gross, *Proc. Natl. Acad. Sci.* **1996**, *93*, 8858.
- [3] C. B. Harley, R. P. Reynolds, *Nucleic Acids Res.* **1987**, *15*, 2343.
- [4] T. Shimada, Y. Yamazaki, K. Tanaka, A. Ishihama, *PLoS One.* **2014**, *9*, 3.

- [5] G. Biffi, D. Tannahill, S. Balasubramanian, *J. Am. Chem. Soc.* **2012**, *134*, 11974.
- [6] E. Y. N. Lam, D. Beraldi, D. Tannahill, S. Balasubramanian, *Nat. Commun.* **2013**, *4*, 1796.
- [7] S. G. Hershman, Q. Chen, J. Y. Lee, M. L. Kozak, P. Yue, L.-S. Wang, F. B. Johnson, *Nucleic Acids Res.* **2007**, *36*, 144.
- [8] A. Verma, K. Halder, R. Halder, V. K. Yadav, P. Rawal, R. K. Thakur, F. Mohd, A. Sharma, S. Chowdhury, *J. Med. Chem.* **2008**, *51*, 5641.
- [9] J. A. Smestad, L. J. Maher, *BMC Med. Genet.* **2015**, *16*, 91.
- [10] I. T. Holder, J. S. Hartig, *Chem. Biol.* **2014**, *21*, 1511; promoter library link: <http://parts.igem.org/Promoters/Catalog/Anderson>
- [11] A. Siddiqui-Jain, C. L. Grand, D. J. Bearss, L. H. Hurley, *Proc. Natl. Acad. Sci.* **2002**, *99*, 11593.
- [12] H. Tateishi-Karimata, N. Isono, N. Sugimoto, *PLoS One.* **2014**, *9*, 3.
- [13] H. Tateishi-Karimata, K. Kawauchi, N. Sugimoto, *J. Am. Chem. Soc.* **2018**, *140*, 642.
- [14] T. Endoh, Y. Kawasaki, N. Sugimoto, *Angew. Chemie Int. Ed.* **2013**, *52*, 5522.
- [15] S. Takahashi, K. T. Kim, P. Podbevšek, J. Plavec, B.H. Kim, N. Sugimoto, *J. Am. Chem. Soc.* **2018**, *140*, 5774.
- [16] Y. Chen, D. Yang, *Curr. Protoc. Nucleic Acid Chem.* **2012**, *17*, 5.
- [17] M. Gajarský, M. L. Živković, P. Stadlbauer, B. Pagano, R. Fiala, J. Amato, L. Tomáška, J. Šponer, J. Plavec, L. Trantírek, *J. Am. Chem. Soc.* **2017**, *139*, 3591.
- [18] V. Limongelli, S. De Tito, L. Cerofolini, M. Fragai, B. Pagano, R. Trotta, S. Cosconati, L. Marinelli, E. Novellino, I. Bertini, A. Randazzo, C. Luchinat, M. Parrinello, *Angew. Chemie Int. Ed.* **2013**, *52*, 2269.
- [19] L. Cerofolini, J. Amato, A. Giachetti, V. Limongelli, E. Novellino, M. Parrinello, M. Fragai, A. Randazzo, C. Luchinat, *Nucleic Acids Res.* **2014**, *42*, 13393.
- [20] S. Takahashi, J. A. Brazier, N. Sugimoto, *Proc. Natl. Acad. Sci.* **2017**, *114*, 9605.
- [21] N. Q. Do, A.T. Phan, *Chem. Eur. J.* **2012**, *18*, 14752.
- [22] V. M. Marathias, P. H. Bolton, *Nucleic Acids Res.* **2000**, *28*, 1969.
- [23] V. T. Mukundan, A. T. Phan, *J. Am. Chem. Soc.* **2013**, *135*, 5017.
- [24] J. Liu, S. Xiao, Y. Hao, Z. Tan, *Angew. Chemie Int. Ed.* **2015**, *54*, 8992.
- [25] V. A. Bloomfield, D. M. Crothers, I. Tinoco Jr., *Nucleic Acids: Structure, Properties, and Functions*, University Science Books, Sausalito, **2000**, p. 90.
- [26] V. Kuryavyi, D. J. Patel, *Structure* **2010**, *18*, 73.
- [27] P. Sket, M. Crnugelj, J. Plavec, *Bioorg. Med. Chem.* **2004**, *12*, 5735.
- [28] T. Agarwal, S. Roy, S. Kumar, T.K. Chakraborty, S. Maiti, *Biochemistry* **2014**, *53*, 3711.
- [29] J. Lopes, A. Piazza, R. Bermejo, B. Kriegsman, A. Colosio, M.-P. Teulade-Fichou, M. Foiani, A. Nicolas, *EMBO J.* **2011**, *30*, 4033.
- [30] M. Marušič, J. Plavec, *Angew. Chemie Int. Ed.* **2015**, *54*, 11716.