



Effect of mutations in a simian virus 40 PolyA signal enhancer on green fluorescent protein reporter gene expression

H.G. Wang^{1,2}, X.F. Wang¹, X.Y. Jing¹, Z. Li³, Y. Zhang¹ and Z.J. Lv¹

¹Hebei Key Lab of Laboratory Animal,
Department of Genetics, Hebei Medical University, Shijiazhuang,
Hebei Province, China

²Department of Biochemistry and Molecular Biology,
Medical College, Henan University, Kaifeng, Henan Province, China

³Department of Respiratory Medicine,
The First Hospital of Hebei Medical University,
Shijiazhuang, Hebei Province, China

Corresponding authors: X.F. Wang / Z.J. Lv
E-mail: wangxiufang@live.cn / lslab@hebmh.edu.cn

Genet. Mol. Res. 10 (3): 1866-1883 (2011)

Received April 11, 2011

Accepted June 19, 2011

Published August 26, 2011

DOI <http://dx.doi.org/10.4238/vol10-3gmr1169>

ABSTRACT. Our previous studies have shown that tandem *Alu* repeats inhibit green fluorescent protein (GFP) gene expression when inserted downstream of the GFP gene in the pEGFP-C1 vector. We found that the 22R sequence (5'-GTGAAAAAATGCTTTATTTGT-3') from the antisense PolyA (240 bp polyadenylation signal) of simian virus 40, eliminated repression of GFP gene expression when inserted between the GFP gene and the *Alu* repeats. The 22R sequence contains an imperfect palindrome; based on RNA structure software prediction, it forms an unstable stem-loop structure, including a loop, a first stem, a bulge, and a second stem. Analysis of mutations of the loop length of the 22R sequence showed that the three-nucleotide loop (wild-type, 22R) induced much stronger GFP expression than did other loop lengths. Two mutations, 4TMI (A7→T, A17→T) and 5AMI (A6→T, T18→A),

which caused the base type changes in the bulge and in the second stem in the 22R sequence, induced stronger GFP gene expression than 22R itself. Mutation of the bulge base (A17→T), leading to complete complementation of the stem, caused weaker GFP gene expression. Sequences without a palindrome (7pieA, 5'-GTGAAAAAATG CAAAAAAGT-3', 7pieT, 5'-GTGTTTTTTTTGCTTTTTTGT-3') did not activate GFP gene expression. We conclude that an imperfect palindrome affects and can increase GFP gene expression.

Key words: Stem-loop; GFP; Alu; SV40 PolyA; Mutation

INTRODUCTION

Noncoding sequences generally exist in the genomes of eukaryotic organisms and contain important genetic information about biological development and physiology such as information about the promoter, enhancer, insulator, and evolution of the sequence. However, the function of most noncoding sequences is unclear (Costantini and Bernardi, 2008; Costa, 2008).

Our previous studies found that single-strand DNA sequences in the human genome, which are mostly noncoding DNA, possessed complementary sequences (palindrome sequences) (Lu et al., 2003). In the human genome, most DNA sequences of certain length (e.g., ~100 bases) can fold into a stem-loop secondary structure. Stem-loop structures are important for mRNA translation, transcription termination, virus replication, and mRNA maturation (Ueno et al., 2009; Cai et al., 2010). Deletion or disruption of predicted stem-loop structures has been demonstrated to affect minus-strand RNA synthesis, and their complementarity in the minus-strand RNA influences the efficiency of plus-strand synthesis (Frolov et al., 2001).

Analyses of mutant promoters in the N4 virion RNA polymerase reveal that DNA secondary structure, specifically a 3-base loop and a 5-7-bp stem hairpin on the template strand, is required for virion RNA polymerase promoter recognition (Dai et al., 1998; Gleghorn et al., 2008). The untranslated regions (UTRs) of the murine hepatitis virus (MHV) contain cis-acting sequences and secondary structures, which are known to have essential roles in RNA synthesis, gene expression, and virion assembly (Kang et al., 2006). Base 35 (A35), located in the stem middle of the 5' UTR first stem-loop (SL1) in the MHV genome, is an extrahelical nucleotide that forms a bulge that destabilizes the SL1 structure. Deleting A35 leads to complete complementation of the stem. Viruses that contain a deletion at A35 have been demonstrated to have strong negative genetic selection, which indicates that the imperfect complementation of the stem is important for viral replication (Li et al., 2008).

Our previous results showed that Alu tandem sequences inhibited green fluorescent protein (GFP) gene expression when inserted downstream of the GFP gene in the pEGFP-C1 vector (Wang et al., 2009a,b). Antisense simian virus 40 (SV40) PolyA signal (PolyAas, 240 bp) eliminated the repression of GFP expression induced by the Alu tandem repeats when placed between the GFP gene and the Alu tandem repetitive sequences. By dividing PolyAas, we found that a second 60-bp sequence, designated 2F2R, activated GFP gene expression (Yin et al., 2010). This finding led us to investigate the mechanisms for eliminating the repression of GFP gene expression.

In this study, 2F2R was divided into different fragments, which were then inserted into the pAlu14 plasmid (14 copies of Alu inserted in the sense orientation downstream of the GFP gene in the pEGFP-C1 vector). The results showed that only the 22R sequence (5'-GTGAAAAAATGCTTTATTTGT-3'), in which there is an imperfect palindrome, significantly activated the GFP gene. To understand how the imperfect palindrome in the 22R sequence activates GFP gene expression, we mutated the 22R sequence in order to increase or decrease the complementarity of the palindrome and found that all mutated sequences that were inserted into the pAlu14 resulted in lower GFP gene expression compared to 22R itself, except 4TMI and 5AMI. Our results suggest that suitably imperfect palindrome may play a role in activating GFP gene expression.

MATERIAL AND METHODS

Construction of expression vector

The pAlu14 plasmid was previously constructed in our laboratory by inserting 14 head-to-tail tandem Alu (283 bp) elements downstream of the GFP gene in the pEGFP-C1 vector (Wang et al., 2009a).

Primers were designed with suitable restriction enzyme sites (*EcoRI/XbaI*; *KpnI/NheI*), and polymerase chain reaction (PCR) was used to amplify the synthetic DNA sequences (as templates) that contained mutated sites or fragments of PolyAas DNA sequences. The PCR products were digested with suitable restriction enzymes and inserted downstream of the GFP gene in pAlu14. When the adhesive ends of the DNA fragments digested with *XbaI* and *NheI* restriction enzymes were ligated by T4 DNA ligase (Fermentas, Lithuania), both recognition sites for *XbaI* and *NheI* were destroyed. Using this characteristic, expression vectors of two inserted tandem sequences were obtained. The primers and templates used for the construction of expression vectors are shown in Tables 1 and 2, respectively.

Cell culture and cell transfection

HeLa cells or MCF7 cells were grown in Dulbecco's modified Eagle's medium (DMEM) with 10% fetal bovine serum. Cells were plated in each well of a 24-well plate at 0.9×10^5 cells/well and grown at 37°C under 5% CO₂ for 30-36 h. At approximately 50-70% confluence, the cells were transiently transfected with 0.4 µg plasmid DNA using 2 µL Lipofectamine™ 2000 reagent (Invitrogen, USA) according to manufacturer instructions and subsequently cultured for an additional 30-36 h. The cells were used for RNA extraction, fluorescence monitoring and flow cytometry determination.

Assessment of GFP expression

Transfected HeLa cells were fixed in 4% paraformaldehyde. Expression of the GFP was assessed by inspecting the HeLa cells transfected with plasmids using a fluorescence microscope (Nikon TE2000-U, Japan). Cell photographs were taken using both light and fluorescence microscopy using the same scope. Transfected HeLa cells were harvested, and GFP expression was assessed by flow cytometry (Beckman Coulter, USA).

Table 1. Primers used for construction of expression vectors.

Names of primers	Sequences of primers	Annotation
Poly60-2F (forward primer)	<i>EcoRI</i> <i>XbaI</i> 5'-ATCGGAATTCCTTAATCTAGATAAGTGAAAAAATGCTTTATT	Amplifying 2F2R
Poly60-2R (reverse primer)	<i>KpnI</i> <i>NheI</i> 5'-ATCGGGTACCATGCTAGCATAATGGTTACAATAAAG	Amplifying 2F2R, Poly4
Poly4F (forward primer)	<i>EcoRI</i> <i>XbaI</i> 5'-ATCGGAATTCCTTAATCTAGATAATGATGCTATTG	Amplifying Poly4
FirLoopF (forward primer)	<i>EcoRI</i> <i>XbaI</i> 5'-ATCGGAATTCCTTAATCTAGATAATGTGAAA AAA	Amplifying 22R, 19R, 16R, 0 nt, 1 nt, 2 nt, 4 nt, 5 nt, 6 nt, VE2A, VE3A
VER (reverse primer)	<i>KpnI</i> <i>NheI</i> 5'-ATCGGGTACCATGCTAGCACAA	Amplifying VE2A, VE3A
Poly22R (reverse primer)	<i>KpnI</i> <i>NheI</i> 5'-ATCGGGTACCATGCTAGCACAAATAAAGCA	Amplifying 22R
SecloopF (forward primer)	<i>EcoRI</i> <i>XbaI</i> 5'-ATCGGAATTCCTTAATCTAGATAATGCTTTATTGTG	Amplifying Secloop
SecloopR (reverse primer)	<i>KpnI</i> <i>NheI</i> 5'-ATCGGGTACCATGCTAGCACAAATTCAC	Amplifying Secloop
1619MR (reverse primer)	<i>KpnI</i> <i>NheI</i> 5'-ATCGGGTACCATGCTAGCAC	Amplifying 19R, 16R
EcoXba (forward primer)	<i>EcoRI</i> <i>XbaI</i> 5'-ATCGGAATTCCTTAATCTAGA	Amplifying 17ntAT
KpnNhe (reverse primer)	<i>KpnI</i> <i>NheI</i> 5'-ATCGGGTACCATGCTAGC	Amplifying 17ntAT
FirLoopR (reverse primer)	<i>KpnI</i> <i>NheI</i> 5'-ATCGGGTACCATGCTAGCACAAATAA	Amplifying 0 nt, 1 nt, 2 nt, 4 nt, 5 nt, 6 nt
22RF (forward primer)	<i>EcoRI</i> <i>XbaI</i> 5'-ATCGGAATTCCTTAATCTAGATAATGTG	Amplifying S2, S4, S5, S6, S7, S8, SeS2, SeS4, SeS5, SeS6, SeS7, SeS8, 4C, 4G, 4T, CG, CC, 1A, 2A, 3A, 4T, 5A, 6A, 7A, 1AMI, 2AMI, 3AMI, 4TMI, 5AMI, 6AMI, 7AMI, 7pieA, 7pieT
22RR (reverse primer)	<i>KpnI</i> <i>NheI</i> 5'-ATCGGGTACCATGCTAGCAC	Amplifying S2, S4, S5, S6, S7, S8, SeS2, SeS4, SeS5, SeS6, SeS7, SeS8, 4C, 4G, 4T, CG, CC, 1A, 2A, 3A, 4T, 5A, 6A, 7A, 1AMI, 2AMI, 3AMI, 4TMI, 5AMI, 6AMI, 7AMI, 7pieA, 7pieT

RT-PCR

Total RNA was extracted from transfected HeLa cells using the TRIzol reagent (Invitrogen). Reverse transcription was carried out using AMV reverse transcriptase (SBS, China) and a 9-nt random primer (SBS). The primers used for PCR were forward, 1GFP1F: 5'-TGAGCCACC GCGCCCAGC-3', and reverse, Alu1R: 5'-TGAGCCACCGCGCCCAGC-3'. The PCR conditions were 94°C for 30 s, 50°C for 30 s, and 72°C for 1 min for 30 cycles in a 25-µL reaction buffer.

Northern blotting

Total RNA from transfected cells was extracted with the TRIzol reagent. RNA was electrophoresed on a 1.2% agarose gel containing 0.4 M formaldehyde, and the RNA was then transferred to nylon membranes (pore diameter = 0.45 µm; Osmonics, USA). A 590-bp fragment from the GFP gene in the pEGFP-C1 plasmid was amplified by PCR using the forward primer 5'-GGGCGAGGGCGATG-3' and reverse primer 5'-CTTGTACAGCTCGTCCATGC-3'. The PCR product was purified by agarose gel electrophoresis and radiolabeled with [α -³²P]-dCTP (Furui, China) using the random primer labeling system (TaKaRa, Japan). The nylon membranes blotted with RNA were hybridized with [α -³²P]-radiolabeled DNA probes at 42°C in 50% formamide-5X SSC (saline sodium citrate)-5X Denhardt's solution-100 µg/mL salmon

sperm DNA for 24 h in a UL2000 hybriLinker (UVP, USA). The membranes were washed using a solution of 1X SSC-0.1% SDS, twice at room temperature, and then washed using a solution of 0.1X SSC-0.1% SDS, three times at 68°C. Autoradiography was then performed. The membranes were subsequently stripped by washing twice at 80°C for 1 h in a solution containing 50% formamide-5% SDS-50 mM Tris, pH 7.4, and then hybridized with an [α - 32 P]-labeled probe for neo gene (the cassette for neomycin resistance) RNA. A 671-bp fragment from the neo gene in the pEGFP-C1 plasmid was amplified by PCR using the forward primer 5'-CACAAACAGACAATCGGCTGCT-3' and the reverse primer 5'-AGCGGCGATACCGTAAAGCAC-3'. The probe for neo RNA was prepared using the random primer labeling system with the 671-bp neo fragment as the template.

Table 2. Synthetic templates used for construction of expression vectors.

Names of templates	Sequences of synthetic templates
19RM	5'- AATGTGAAAAAAAAATGCTTTATTGCTAGC
16RM	5'- AATGTGAAAAAAAAATGCTTTGCTAGC
17ntAT	5'- CTAGATAATAAAAAAAAAATGCTTTATTGCTAGCAT
Loop0nt	5'- GTGAAAAAAAAATTATTAGT
Loop1nt	5'- GTGAAAAAAAAGTTTATTGT
Loop2nt	5'- GTGAAAAAAAAATGTTTATTGT
Loop4nt	5'- GTGAAAAAAAAACTGCTTTATTGT
Loop5nt	5'- GTGAAAAAAAAACGTGCTTTATTGT
Loop6nt	5'- GTGAAAAAAAAATCGTCTTTATTGT
S2	5'- ATAATGTGAAAAAAAAATGCTTTATTGCTAGC
S4	5'- ATAATGTGAAAAAAAAATGCTTTATTGCTAGC
S5	5'- ATAATGTGAAAAAAAAATGCTTTTATTGCTAGC
S6	5'- ATAATGTGAAAAAAAAATGCTTTTTATTGCTAGC
S7	5'- ATAATGTGAAAAAAAAATGCTTTTTATTGCTAGC
S8	5'- ATAATGTGAAAAAAAAATGCTTTTTATTGCTAGC
SeS2	5'- ATAATGTGAAAAAAAAATGCTTTATTGCTAGC
SeS4	5'- ATAATGTGAAAAAAAAATGCTTTATTGCTAGC
SeS5	5'- ATAATGTGAAAAAAAAATGCTTTTATTGCTAGC
SeS6	5'- ATAATGTGAAAAAAAAATGCTTTATTGCTAGC
SeS7	5'- ATAATGTGAAAAAAAAATGCTTTTATTGCTAGC
SeS8	5'- ATAATGTGAAAAAAAAATGCTTTTTATTGCTAGC
VE2A	5'- AATGTGAAAAAAAAATGCTTAAATTGTGCTA
VE3A	5'- AATGTGAAAAAAAAATGCTTAAATTGTGCTA
4C	5'- AATGTGAAAAAAAAATGCTTTCTTTGTGCTA
4G	5'- AATGTGAAAAAAAAATGCTTTGTTGTGCTA
4T	5'- AATGTGAAAAAAAAATGCTTTTTGTGCTA
CG	5'- AATGTGAAACAAATGCTTTGTTGTGCTA
CC	5'- AATGTGAAACAAATGCTTTCTTTGTGCTA
4TMI	5'- AATGTGAAATAAATGCTTTTTGTGCTA
1A	5'- AATGTGAAAAAAAAATGCATTATTGTGCTA
2A	5'- AATGTGAAAAAAAAATGCTATATTGTGCTA
3A	5'- AATGTGAAAAAAAAATGCTTAAATTGTGCTA
4T	5'- AATGTGAAAAAAAAATGCTTTTTGTGCTA
5A	5'- AATGTGAAAAAAAAATGCTTAAATTGTGCTA
6A	5'- AATGTGAAAAAAAAATGCTTATATTGTGCTA
7A	5'- AATGTGAAAAAAAAATGCTTTATTAGTGCTA
1AMI	5'- AATGTGAAAAAATGCATTATTGTGCTA
2AMI	5'- AATGTGAAAAAATGCTATATTGTGCTA
3AMI	5'- AATGTGAAAAAATGCTTAAATTGTGCTA
4TMI	5'- AATGTGAAAAAATGCTTTTTGTGCTA
5AMI	5'- AATGTGAAAAAATGCTTAAATTGTGCTA
6AMI	5'- AATGTGATAAAAAATGCTTTATATTGTGCTA
7AMI	5'- AATGTGTAAAAAATGCTTTATTAGTGCTA
7pieA	5'- AATGTGAAAAAATGCAAAAAAGTGCTA
7pieT	5'- AATGTGTTTTTTTGTCTTTTTTGTGCTA

RNA degradation analysis

HeLa cells were transfected with p4T*2-Alu14, p1AMI*2-Alu14, p2AMI*2-Alu14, or p4TMI*2-Alu14 expression vectors. Thirty hours later, actinomycin D (5 µg/mL) was added to block RNA synthesis. RNA was extracted after 0-, 2- or 4-h incubation and Northern blotting was performed.

RESULTS

Dividing 2F2R and searching for the critical sequence that activates GFP gene expression

Our previous results showed that the 60-bp sequence 2F2R from PolyAas upregulated GFP gene expression (Yin et al., 2010). It has been reported that the DNA regions prone to strand separation and hairpin formation play a key role in regulating gene expression (Xu et al., 2001). To determine which fragment is critical for activating GFP gene expression, the 2F2R sequence was divided into 3 fragments (22R, Secloop and Poly4; Figure 1A). The 22R sequence is a fragment from 1 to 22 bp in 2F2R; Secloop is a fragment from 11 to 32 bp in 2F2R, and Poly4 is a fragment from 31 to 60 bp in 2F2R. Two tandem repeats of these fragments were inserted between the GFP gene and the Alu tandem repeats in the pAlu14 plasmid (Figure 1B), and HeLa cells were transfected with these expression vectors. Northern blot analysis showed that only the 22R significantly eliminated the inhibition of GFP gene expression induced by Alu14 (Figure 1C and 1D, lane 1 vs lane 4).

The 22R sequence contains a palindrome so that it is possible for 22R to form stem-loop structures. One of the stem-loop structures that may form in the 22R sequence included a 3-nt loop (TGC), a 3-bp first stem, a 2-nt bulge, and a 3-bp second stem (Figure 1F) on the basis of structure predictions by the RNA Structure 4.6 software and the principle of base pairing. To determine the effects of the 5' and 3' bases in the 22R sequence on gene expression, several bases at the 5' and/or 3' terminal in 22R were deleted. GFP gene expression decreased significantly when 3 nt (TGT) were deleted from the 22R 3' end (19R; Figure 1A) (Figure 1D, lane 2 vs lane 1; Figure 1E, lane 3 vs lane 4), which indicates that these 3 nt are important for activating GFP gene expression. GFP gene expression also significantly decreased when 6 nt were deleted from the 3' end of 22R (16R; Figure 1A) (Figure 1D, lane 3 vs lane 1; Figure 1E, lane 1 vs lane 4). The level of GFP gene expression was not obviously changed when 3 nt (GTG) from the 5' end and 2 nt (GT) from the 3' end of 22R were deleted (17ntAT; Figure 1A) (Figure 1E, lane 2 vs lane 4). We conclude that 17ntAT is a critical sequence for activating GFP gene expression. Because insertion of 22R in the pAlu14 plasmid induced the production of a higher molecular mass GFP RNA (indicated by arrow in Figure 1D, lane 1), we used 22R as the fundamental experimental sequence to study the effects of base mutation on GFP gene expression.

The effects of the base number in the 22R loop and stems on GFP gene expression

To determine the role of the number of bases in the 22R loop and stems on GFP gene expression, we changed the base number within the loop and stems (Figure 2B and D). Chang-

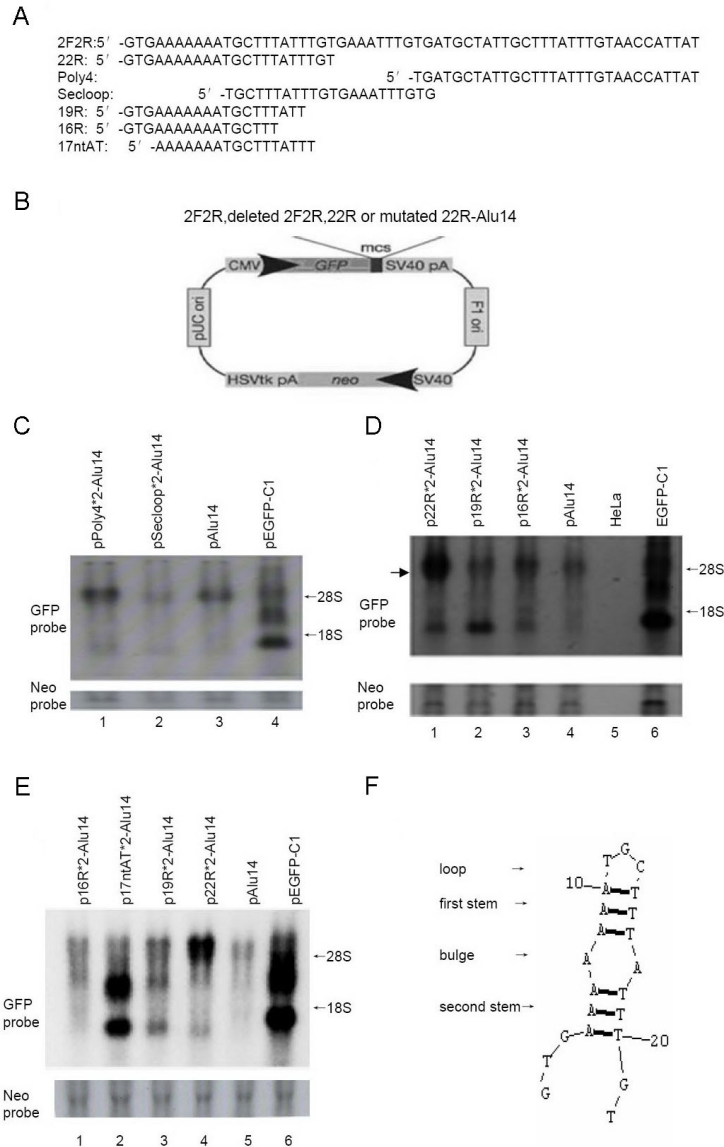


Figure 1. Dividing 2F2R and searching for the critical sequence that activates green fluorescent protein (GFP) gene expression. **A.** 2F2R sequence was deleted to form 22R, Poly4 and Secloop. 19R, 16R and 17ntAT were obtained by deleting specific bases from 22R. **B.** The inserted position of DNA segments in pEGFP-C1 vector. **C. D. E.** The two tandem sequences of Poly4, Secloop, 22R, 19R, 16R, or 17ntAT were inserted downstream of the GFP gene in pAlu14. HeLa cells were transfected with these expression vectors, and GFP RNA was detected by Northern blot analysis. **F.** The 22R DNA sequence that contains an imperfect palindrome is predicted to form an unstable stem-loop structure including a loop (3 nt), the first stem (3 bp), a bulge (2 nt), and the second stem (3 bp). There are some sites for transcription termination in these expression vectors used, so that in most blots, more than one band was visible after reaction with the GFP-specific probe. We judged the amount of GFP transcripts in the blots on the basis of clearly visible bands. The positions for 28S and 18S rRNA were obtained by methylene blue staining.

ing the loop base number from 0 to 6 nt showed that only the sequence containing a 3-nt loop (wild-type) significantly activated GFP gene expression (Figure 2A, lane 4 vs lane 8), whereas other mutants showed only weak activation (Figure 2A, lanes 1-3 and 5-7 vs lane 8). The base pair number of the first stem was changed to 2, 4, 5, 6, 7, or 8 bp, and the results showed that only the sequence containing 3 bp (wild-type) significantly activated GFP gene expression (Figure 2C, lane 13 vs lanes 1-6 and lane 14). Changing the second stem to 2, 4, 5, 6, 7, and 8 bp showed that the sequence containing 3 bp (wild-type, 22R) significantly activated gene expression (Figure 2C, lane 13 vs lanes 7-12 and lane 14).

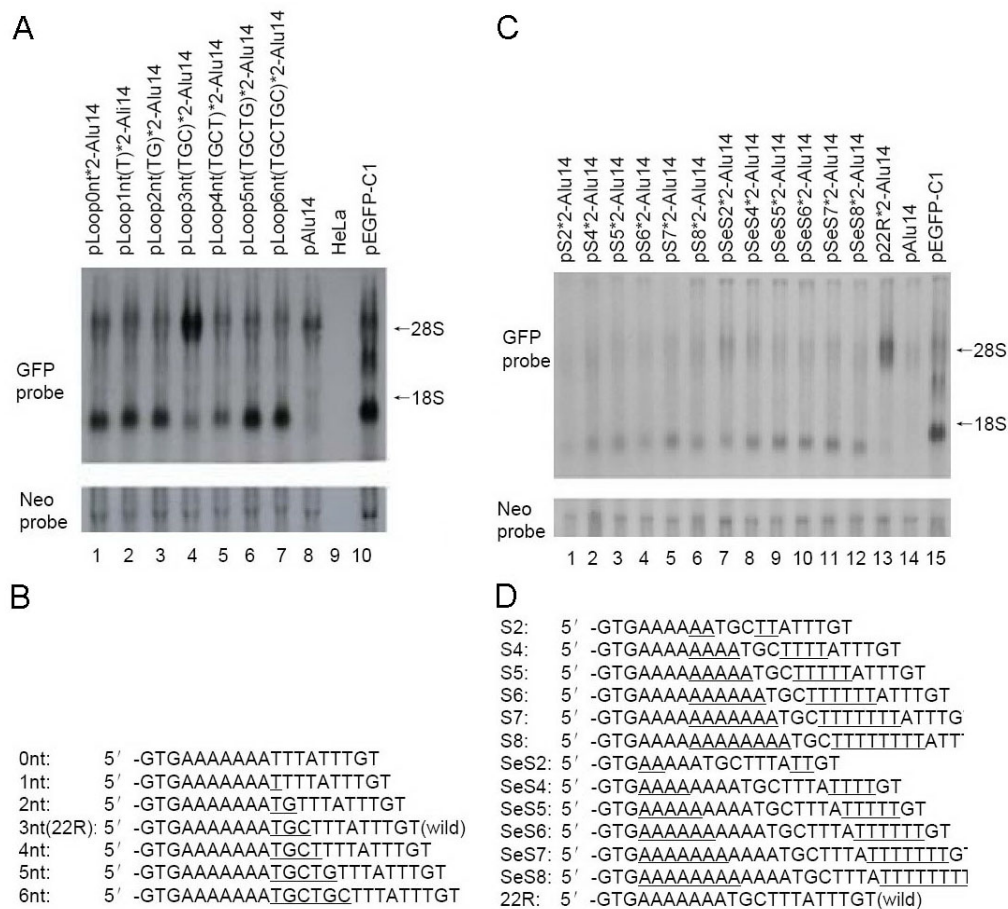


Figure 2. The effects of the base number in the 22R loop and stems on green fluorescent protein (GFP) gene expression. **A.** The base number of the 22R loop was changed to 0, 1, 2, 4, 5, or 6 nt. Two tandem sequences were inserted downstream of the GFP gene in pAlu14 to construct expression vectors. HeLa cells were transfected with these expression vectors, and GFP RNA was detected by Northern blot analysis. **B.** The sequences of the loop mutants. The loop base position is underlined. **C.** The base pair number of the 22R first stem was changed to 2 (S2), 4 (S4), 5 (S5), 6 (S6), 7 (S7), or 8 bp (S8) (3 bp, wild-type). The base pair number of the 22R second stem was changed to 2 (SeS2), 4 (SeS4), 5 (SeS5), 6 (SeS6), 7 (SeS7), or 8 bp (SeS8) (3 bp, wild-type). Two tandem sequences were inserted downstream of the GFP gene in pAlu14 to construct expression vectors. HeLa cells were transfected with these expression vectors, and GFP RNA was detected by Northern blot analysis. **D.** The sequences of the first and second stem mutants. Stem bases are underlined.

The effect of change of base type and length in the 22R bulge on GFP gene expression

To examine the effect of the size of the bulge, the base numbers of the 22R bulge were changed to 4 nt (VE2A, Figure 3B) or 6 nt (VE3A, Figure 3B). The results showed that the 22R (2 nt bulge) induced much stronger GFP gene expression than the mutants containing 4 or 6 nt in the bulge (Figure 3A, lane 1 vs lanes 2, 3). This shows that the base number of the bulge in 22R plays an important role on GFP gene expression. Mutation of the bulge base type in 22R showed that only the mutant containing a T/T (4TMI, Figure 3D) bulge enhanced GFP gene expression (Figure 3C, lane 6 vs lane 8). Furthermore, the 4TMI mutant induced much stronger GFP expression compared to 22R (Figure 3C, lane 6 vs lane 7).

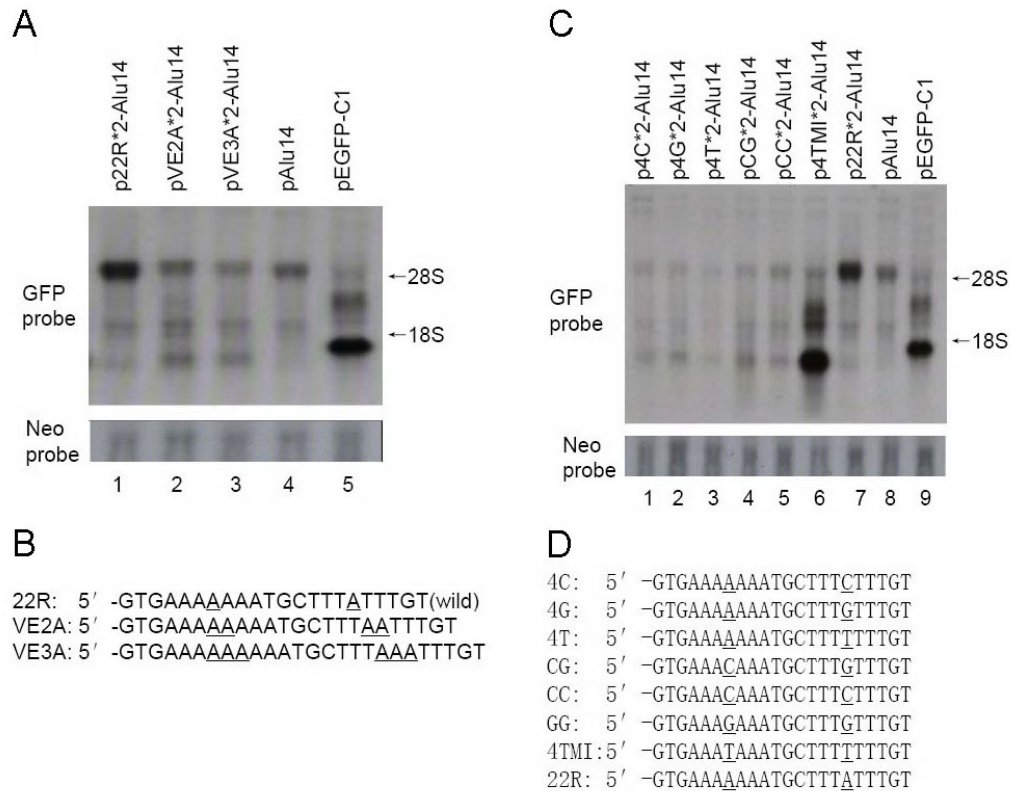


Figure 3. The effect of change of base type and length in the 22R bulge on green fluorescent protein (GFP) gene expression. **A.** The base number of the 22R bulge was changed to 4 (VE2A) or 6 nt (VE3A) (2 nt, wild-type). Two tandem sequences were inserted downstream of the GFP gene in the pAlu14 vector. HeLa cells were transfected with these expression vectors, and GFP RNA was detected by Northern blot analysis. **B.** The sequences of the bulge mutants. Bulge bases are underlined. **C.** The bases of the 22R bulge were changed to A/C (4C), A/G (4G), A/T (4T), C/G (CG), C/C (CC), or T/T (4TMI) (A/A, wild-type). Two tandem sequences were inserted downstream of the GFP gene in pAlu14 to construct the expression vectors. HeLa cells were transfected with these expression vectors, and GFP RNA was detected by Northern blot analysis. **D.** Sequences of bulge mutants. Bulge bases are underlined.

The effects of specific bases in 22R on GFP gene expression

To examine the effects of specific bases in 22R on GFP gene expression, the 22R sequence was mutated at single bases; the T bases of the first stem and the second stem in 22R were mutated to A, and A in the bulge (A17) was mutated to T. We hypothesized that point mutations in the stems would decrease the stability of the stem-loop structure, whereas a point mutation in the bulge, changing A17 to T (4T, Figure 4D), would stabilize the stem-loop structure. To study further the effects of stem-loop structure stability on GFP gene expression, stem complementation and bulge were restored. Sequences of single base mutations and restored complementation mutations are shown in Figure 4D. The results of single base mutations in the 22R stems and bulge (A17→T) showed that all mutants resulted in weaker GFP gene expression compared to 22R itself (Figure 4A, lanes 1-4 vs lane 9; Figure 4B, lanes 1-3 vs lane 7). Among the mutants of restored bulge and stem complementation, the 4TMI (Figure 4A, lane 8 vs lane 9) and 5AMI (Figure 4B, lane 4 vs lane 7) mutants caused much stronger GFP expression compared to 22R itself, whereas five other restored complementation mutants induced much less GFP expression (Figure 4A, lanes 4-7 vs lane 9; Figure 4B, lanes 5, 6 vs lane 7). To confirm that 4TMI (Figure 4A, lane 8) induced much stronger GFP gene expression than did 22R, GFP RNA levels of HeLa cells transfected with p4T*2-Alu14, p4TMI*2-Alu14, p22R*2-Alu14, or pAlu14 were detected using RT-PCR. Figure 4E shows that the 4T mutant caused weaker GFP gene transcription compared to 22R (Figure 4E, lane 1 vs lane 3), whereas the 4TMI mutant induced stronger GFP gene transcription (Figure 4E, lane 2 vs lane 3).

Figure 4A and B show that the 4TMI and the 5AMI induced stronger GFP gene expression compared to 22R, whereas the 3AMI caused weaker GFP gene expression in HeLa cells. The results indicate that the GFP gene was activated when the bulge base was mutated to T/T (4TMI) from A/A in 22R and that it caused different results when bulge (A7/A17) peripheral base pairs were mutated to T/A from A/T in 22R. To confirm whether this phenomenon exists in other cell lines, we transfected the expression vectors into MCF7 cells. Figure 4C shows that the 4TMI and the 5AMI induced stronger GFP expression compared to 22R (Figure 4C, lanes 2, 3 vs lane 4), whereas the 3AMI caused weaker GFP expression (Figure 4C, lane 1 vs lane 4) in MCF7 cells, which is consistent with the results in HeLa cells.

The 4TMI and 5AMI mutants generated shorter RNA transcripts than did 22R (indicated by arrow in Figure 4A, lane 8 and Figure 4B, lane 4). All mutants used in this study induced higher levels of GFP RNA transcription than relative to pAlu14 (Figure 4A, B).

We chose p4T*2-Alu14, p5A*2-Alu14, p4TMI*2-Alu14, p5AMI*2-Alu14, p22R*2-Alu14, pAlu14, and pEGFP-C1 to detect GFP expression using flow cytometry (Figure 5A) and fluorescence microscopy (Figure 5B). The results showed that the 4T and 5A mutants caused weaker GFP expression compared to 22R. In contrast, the 4TMI and 5AMI mutants induced much stronger GFP expression than with 22R, which is consistent with the RT-PCR results.

Sequences lacking the palindrome did not activate GFP gene transcription

Palindrome sequences do not exist in 7pieA and 7pieT (Figure 6B). Two tandem 7pieA and 7pieT sequences inserted downstream of the GFP gene in pAlu14 failed to activate GFP gene expression (Figure 6A, lanes 4, 5). The 4T sequence that contains a perfect palindrome also did not activate GFP gene expression (Figure 6A, lane 2; also see Figure 4A, lane 4).

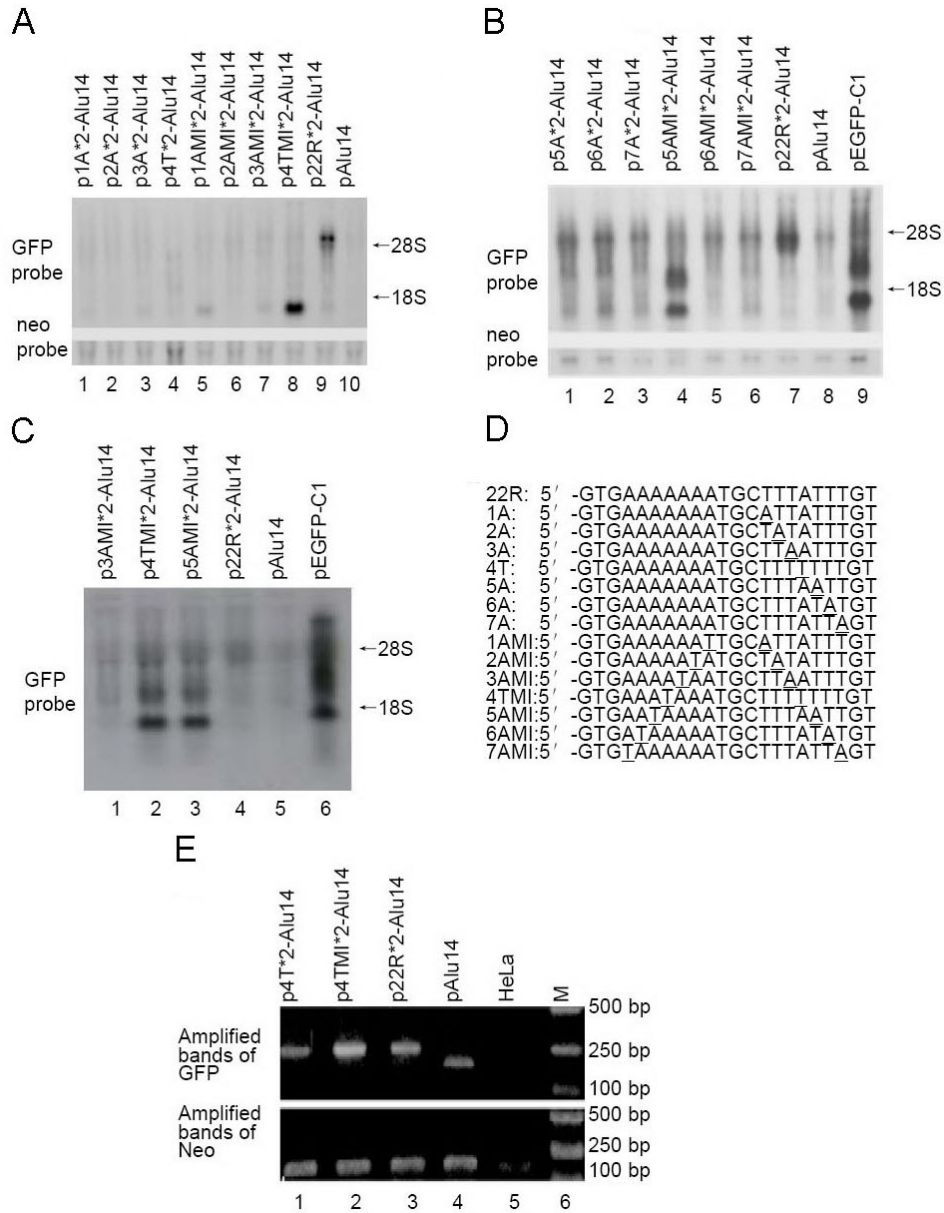


Figure 4. The effects of specific bases in 22R on green fluorescent protein (GFP) gene expression. **A. B.** The T of the first stem and the second stem in 22R was mutated to A. Complementation was restored in mutated stems (e.g., 1AMI). Two tandem sequences were inserted downstream of the GFP gene in pAlu14 to construct the expression vectors. HeLa cells were transfected with these expression vectors, and GFP RNA was detected by Northern blot analysis. **C.** MCF7 cells were transfected with expression vectors of p3AMI*2-Alu14, p4TMI*2-Alu14, p5AMI*2-Alu14, p22R*2-Alu14, pAlu14, and pEGFP-C1. GFP RNA was detected by Northern blot analysis. **D.** The sequences of single base mutations (1A-7A) and the sequences of restored complementation mutations (1AMI-7AMI). Mutated bases are underlined. **E.** RT-PCR results of GFP RNA of 4T, 4TMI, 22R, Alu14, and HeLa.

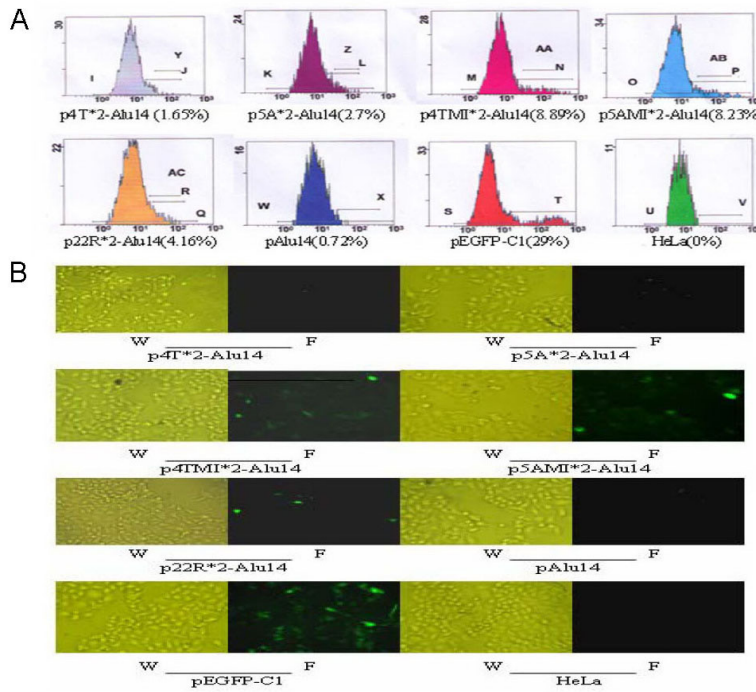


Figure 5. The effects of specific bases in 22R on green fluorescent protein (GFP) expression. HeLa cells were transfected by constructed expression vectors and GFP was detected by **A.** flow cytometry and **B.** fluorescence microscopy.

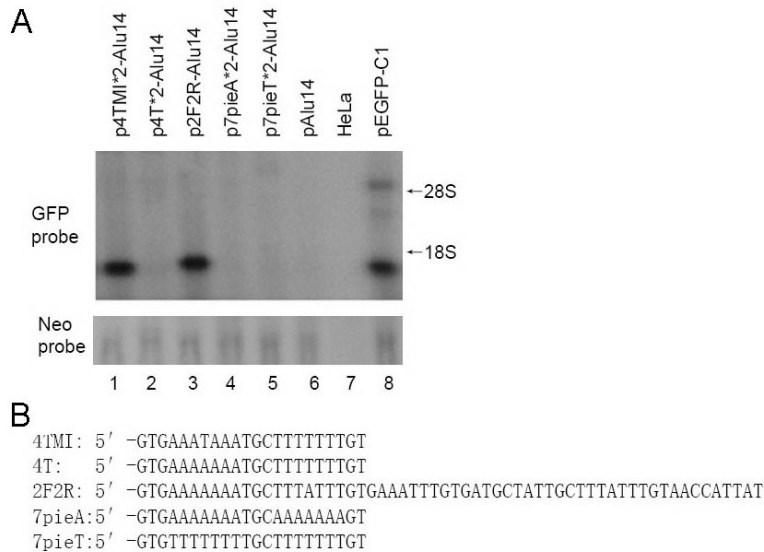


Figure 6. Sequences lacking the palindrome failed to activate green fluorescent protein (GFP) gene expression. **A.** Two tandem sequences of 7pieA and 7pieT were inserted downstream of the GFP gene in pAlu14. HeLa cells were transfected with the expression vectors, and GFP RNA was detected by Northern blot analysis. **B.** Sequences of 4TMI, 4T, 2F2R, 7pieA, and 7pieT.

RNA degradation analysis

To demonstrate that the changes in GFP RNA levels were not due to changes in the rate of mRNA degradation, HeLa cells were transfected with 4 expression vectors (p4T*2-Alu14, p1AMI*2-Alu14, p2AMI*2-Alu14, or p4TMI*2-Alu14) and then treated with actinomycin D. After 0, 2 and 4 h, RNA levels expressed by the different expression vectors were detected (Figure 7A). The relative levels of RNA at 0, 2 and 4 h in p4T*2-Alu14, p1AMI*2-Alu14, p2AMI*2-Alu14, and p4TMI*2-Alu14 were 1, 0.538, 0.380; 1, 0.532, 0.366; 1, 0.549, 0.382, and 1, 0.522, 0.339 (based on data shown in Figure 7B), respectively, by gray scale scan analysis with the Bandscan 5.0 software. We conclude that the degradation rates of RNA from different expression vectors were essentially identical.

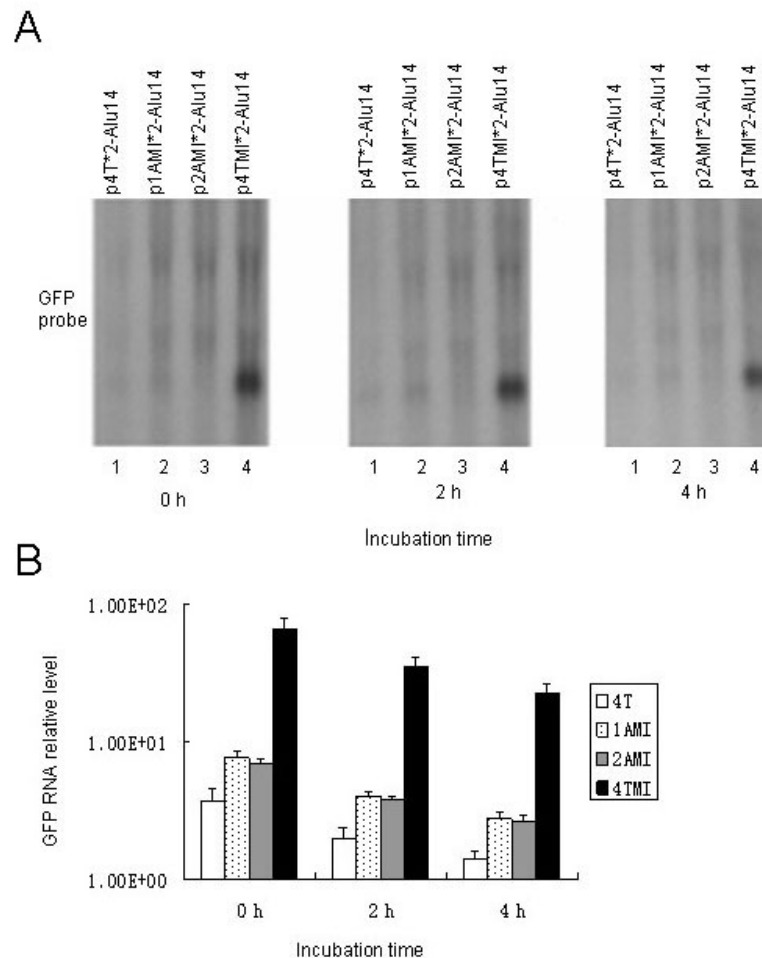


Figure 7. RNA degradation assay. **A.** HeLa cells were transfected with the expression vectors, treated with actinomycin D to inhibit RNA synthesis, and incubated for 0, 2 or 4 h. Northern blot analysis was performed. **B.** Relative RNA levels were calculated from the data presented in Figure 7A (mean \pm SD, N = 3).

DISCUSSION

It has been reported that SV40 PolyA upregulates luciferase gene expression in HeLa cells (Xu et al., 2001). Our studies demonstrate that SV40 PolyAas and its segment 2F2R eliminate the inhibition of GFP gene expression induced by Alu repeats (Yin et al., 2010).

In order to determine which fragments activate the GFP gene in the 2F2R sequence, we divided 2F2R (the second 60 bp of SV40 PolyAas) into three fragments: 22R (1-22 nt), Secloop (11-32 nt) and Poly4 (31-60 nt). Two tandem repeats were inserted between the GFP gene and the Alu tandem repeats in the pAlu14 plasmid. The results showed that only 22R significantly eliminated the inhibition of GFP gene expression induced by Alu repeats.

More than one million copies of the Alu element are interspersed throughout the human genome (Lander et al., 2001; Grimwood et al., 2004). The distribution of Alu elements is not random and there are many clusters of Alus in the human genome that are similar to the Alu tandems used in our study, particularly in human chromosome 19. It is useful to know the blocking and deblocking gene expression properties of Alu repeats for understanding the effects of Alu elements.

Although the 22R sequence was derived from SV40 PolyAas, it is extensively distributed in the human genome. We have searched the human genome by the BLAST software in NCBI and found that there was a lot of 22R and its homologous sequences (data not shown). It is impossible that the eukaryotic genome contains special mechanisms for the expression of virus genes, so the elements of the virus must simulate the process that takes place in eukaryotic cells. We suggest that 22R and its homologous sequences play a role in the regulation of gene expression in eukaryotes.

GFP gene expression significantly decreased when 3 nt were deleted from the 22R 3' end (19R; Figure 1D), indicating that these 3 nt (5'-TGT) are important for activating GFP gene expression (Figure 1D, lane 2 vs lane 1). GFP gene expression also significantly decreased when 6 nt were deleted from the 22R 3' end (16R; Figure 1D, lane 3 vs lane 1). The level of GFP gene expression was not significantly changed when 3 nt (GTG) from the 5' end and 2 nt (GT) from the 3' end of the 22R sequence were deleted (17ntAT; Figure 1E, lane 2 vs lane 4). We conclude that 17ntAT is a critical sequence for activating GFP gene expression.

The stem-loop structure was thought to regulate gene expression (Nelson and Green, 2005). The 5' and 3' UTRs have been demonstrated to be critical in controlling the biological processes of the virus. The 5' and 3' UTRs of viral genomes are predicted to be highly structured. Bio-software predicts various stem-loop structures within these regions. Most stem-loop structures existing in viral genomes show bulge sequences in their stems (Yu and Markoff, 2005; Nickens and Hardy, 2008; Roskopf et al., 2010). The 22R sequence contained an imperfect palindrome and maybe formed an unstable stem-loop structure including a loop (3 nt), the first stem (3 bp), a bulge (2 nt), and the second stem (3 bp) (Figure 1F). To study the effects of the imperfect palindrome on GFP gene expression, the 22R sequence was extensively mutated.

Here, the lengths of the loop, the first stem, the second stem, and the bulge in the 22R sequence were changed; the bases of the bulge and stems in the 22R sequence were changed; mutations that restored complementarity were made, and sequences lacking stem-loop structure were generated (7pieA, 7pieT; Figure 6). Two tandem sequences were inserted downstream of the GFP gene in pAlu14 to construct expression vectors. HeLa cells were transfected with these expression vectors, and GFP RNA was detected. The results showed that: 1) the

4TMI (Figure 3C, lane 6) and 5AMI (Figure 4B, lane 4) mutants induced much stronger GFP gene expression compared to 22R; 2) changing the loop and stem lengths induced weaker GFP expression compared to 22R (Figure 2); 3) changing the bulge length or bases (except 4TMI) induced weaker GFP expression compared to 22R (Figure 3); 4) mutation of single bases in the first and the second stems caused weaker GFP gene expression compared to 22R (Figure 4); 5) restoring complementation of the stem construction did not increase GFP gene expression (except 5AMI) (Figure 4), 6) GFP gene was activated when bulge base was mutated to T/T (4TMI) but repressed when bulge base was mutated to A/T (4T, perfect palindrome), and it caused different results when bulge (A7/A17) peripheral base pairs were mutated to T/A; this phenomenon exists in both HeLa and MCF7 cells (Figure 4), and 7), where sequences lacking a palindrome did not activate GFP gene expression (Figure 6).

DNA cruciform structures form when intrastrand pairing occurs between complementary bases of inverted repeat sequences in double-stranded DNA. Cruciform formation is less energetically favorable than B-form DNA; therefore, the extrusion of such structures from duplex DNA requires the driving energy provided by negative supercoiling (Sean et al., 2009). Hairpin structures in the cruciform promoter for the bacteriophage N4 virion RNA polymerase are extruded at physiological superhelical density (Chou et al., 1999). The palindromes in double-stranded DNA of cells may form stem-loop structures on a small scale (Darlow and Leach, 1998). Therefore, it has been suggested that the structures formed by these palindrome sequences may play an important role in regulating gene expression in cells.

In changing the length of the loop from 0 to 6 nt in the 22R sequence, we found that the wild-type 22R (3 nt) significantly activated gene expression. Having a loop size of 3-4 bases is most optimal for hairpin stability. By either increasing or decreasing the number of bases within the loop, hairpin stability decreases (Kuznetsov et al., 2008). The amount of folding of single-strand DNA hairpins is related to the length of the loop (Li et al., 2009). The ability of the 3-nt loop in 22R to activate the GFP gene may be associated with the stability and folding of this loop.

Mutants of the first stem and the second stem induced weaker GFP gene expression compared to 22R. The stability of stem-loop structures decreases or increases when stem lengths are more than 3 bp or less than 3 bp, which deviates from the most beneficial condition for activating GFP gene. Likewise, when the number of bases associated with the bulge increased, the stability of the stem-loop structure decreased and ability of activating the GFP gene decreased (Figure 3A, lanes 2, 3 vs lane 1).

1AMI, 2AMI, 3AMI, 6AMI, and 7AMI induced weaker GFP gene expression. The ability of forming a variety of stem-loop structures decreased in these sequences. We propose that it is essential to form different types of stem-loop structures to activate the gene. *In vitro* studies have demonstrated that certain DNA sequences can form many types of stem-loop conformations (Li et al., 2009), which supports our hypothesis that the formation of different types of stem-loop structures can activate gene expression.

The A base at position 35 in SL1 (stem-loop 1) of the MHV genome is not complementary (bulge). If A is deleted, the virus cannot replicate, demonstrating that viral replication needs an unstable stem-loop structure (Li et al., 2008). A perfect palindrome (4T) did not enhance GFP gene transcription, and an imperfect palindrome (22R) activated GFP gene expression, which is consistent with the findings in MHV. The structure of the bulge region in 5AMI is similar to the wild-type A35 region in MHV SL1, as both bulge regions contain two A bases, and only one A is complementary.

In this study, the neo gene was used as a control for transfection efficiency of GFP. Neo and GFP exist on the same expression vectors, eliminating the possibility that transfection efficiency influenced the differences observed in the experimental results. To eliminate the possibility that the RNA stability of the GFP transcript caused the differences observed in the experimental results, an RNA degradation experiment was performed. The results showed that with incubation time, GFP RNA in p4T*2-Alu14, p1AMI*2-Alu14, p2AMI*2-Alu14, and p4TMI *2-Alu14 degraded at essentially the same rates.

We examined the relationship between transcription termination and transcription amount. 22R induced higher molecular mass GFP RNA, and loop0nt-loop6nt induced lower molecular mass RNA (Figure 2A); however, the level of transcription caused by 22R was higher than that of loop0nt-loop6nt (Figure 2A). 5AMI caused lower molecular mass GFP RNA, and the level of transcription caused by 5AMI was higher than that of 5A, which induced higher molecular mass transcripts (Figure 4B). These results illustrate that there is no evident relationship between transcription termination and transcription amount.

In this study, we found that mutation of the 22R sequence remarkably influenced GFP gene expression by observing the effects of base change in the 22R sequence on the activation of the GFP reporter gene. There may be two explanations for the mechanisms by which the mutation of the 22R sequence affects gene expression: one is that the sequence itself plays a role by binding a trans-regulatory factor rather than secondary structure; another is that the sequence mutations affect DNA secondary structure (stem-loop structure) that then influences gene expression. Our suggestion that unstable DNA secondary structures may activate gene expression was based on the following data:

- 1) 17ntAT activates GFP gene expression, but 19R cannot activate GFP gene. 17ntAT is the smallest sequence that retains structure;
- 2) 17ntAT activates the GFP gene, but SeS4-SeS8 sequences only induce weaker GFP gene expression. SeS4-SeS8 include 17ntAT sequence;
- 3) The structure of 4TMI is similar with 22R;
- 4) 7pieA and 7pieT cannot activate the GFP gene because they do not contain a palindrome;
- 5) 4T contains a complete palindrome (5'-AAAAAATGCTTTTTTTT-3') and cannot activate the GFP gene.

The results in this study indicate that a suitably imperfect palindrome plays a role in activating the GFP gene. Why did 22R and 4TMI (containing an imperfect palindrome) activate GFP gene expression, whereas 4T (containing a perfect palindrome) and 7pieA and 7pieT (without a palindrome) did not? We propose that the sequences that contain suitably imperfect palindromes activate GFP gene expression by DNA dynamic changes between unstable stem-loop structures and double-strand forms, which is an illative hypothesis that demands more studies for support.

The functions *in vivo* of most noncoding sequences are unknown. The main purpose of this study was to explore new mechanisms by which noncoding sequences regulate gene expression. Suitably imperfect palindromes can activate gene expression, which may be a new mechanism by which noncoding sequences regulate gene expression. Moreover, this study has application value in the biotechnology field. The Alu tandem in pAlu14 is a simple repetitive sequence, so it is easy to explain how repetitive sequences influence gene expression. 22R is a relatively simple sequence, and new research on trans-regulatory factors and other mutations

of 22R could produce new outcomes. In addition, research on the ability of unstable stem-loop structures to activate gene expression may bring about further studies.

ACKNOWLEDGMENTS

Research supported by grants from the Hebei Province Natural Science Foundation of China (#C2008001065 and #C2011206043) and the Key Project of Hebei Province (#08276101D-90).

REFERENCES

- Cai L, Fritz D, Stefanovic L and Stefanovic B (2010). Binding of LARP6 to the conserved 5' stem-loop regulates translation of mRNAs encoding type I collagen. *J. Mol. Biol.* 395: 309-326.
- Chou SH, Tseng YY and Chu BY (1999). Stable formation of a pyrimidine-rich loop hairpin in a cruciform promoter. *J. Mol. Biol.* 292: 309-320.
- Costa FF (2008). Non-coding RNAs, epigenetics and complexity. *Gene* 410: 9-17.
- Costantini M and Bernardi G (2008). Correlations between coding and contiguous non-coding sequences in isochore families from vertebrate genomes. *Gene* 410: 241-248.
- Dai X, Kloster M and Rothman-Denes LB (1998). Sequence-dependent extrusion of a small DNA hairpin at the N4 virion RNA polymerase promoters. *J. Mol. Biol.* 283: 43-58.
- Darlow JM and Leach DR (1998). Evidence for two preferred hairpin folding patterns in d(CGG).d(CCG) repeat tracts *in vivo*. *J. Mol. Biol.* 275: 17-23.
- Frolov I, Hardy R and Rice CM (2001). *Cis*-acting RNA elements at the 5' end of Sindbis virus genome RNA regulate minus- and plus-strand RNA synthesis. *RNA* 7: 1638-1651.
- Gleghorn ML, Davydova EK, Rothman-Denes LB and Murakami KS (2008). Structural basis for DNA-hairpin promoter recognition by the bacteriophage N4 virion RNA polymerase. *Mol. Cell* 32: 707-717.
- Grimwood J, Gordon LA, Olsen A, Terry A, et al. (2004). The DNA sequence and biology of human chromosome 19. *Nature* 428: 529-535.
- Kang H, Feng M, Schroeder ME, Giedroc DP, et al. (2006). Putative *cis*-acting stem-loops in the 5' untranslated region of the severe acute respiratory syndrome coronavirus can substitute for their mouse hepatitis virus counterparts. *J. Virol.* 80: 10600-10614.
- Kuznetsov SV, Ren CC, Woodson SA and Ansari A (2008). Loop dependence of the stability and dynamics of nucleic acid hairpins. *Nucleic Acids Res.* 36: 1098-1112.
- Lander ES, Linton LM, Birren B, Nusbaum C, et al. (2001). Initial sequencing and analysis of the human genome. *Nature* 409: 860-921.
- Li L, Kang H, Liu P, Makkinje N, et al. (2008). Structural lability in stem-loop 1 drives a 5' UTR-3' UTR interaction in coronavirus replication. *J. Mol. Biol.* 377: 790-803.
- Li Y, Ho ES, Gunderson SI and Kiledjian M (2009). Mutational analysis of a Dcp2-binding element reveals general enhancement of decapping by 5'-end stem-loop structures. *Nucleic Acids Res.* 37: 2227-2237.
- Lu ZJ, Zhai Y, Wang XF and Song SX (2003). DNA sequence composition on human X chromosome differing from that on chromosomes 6,7,8,10,11 and 12. *Genet. Sin.* 30: 1051-1060.
- Nelson MJ and Green BR (2005). Double hairpin elements and tandem repeats in the non-coding region of *Adenoides eludens* chloroplast gene minicircles. *Gene* 358: 102-110.
- Nickens DG and Hardy RW (2008). Structural and functional analyses of stem-loop 1 of the Sindbis virus genome. *Virology* 370: 158-172.
- Roskopf JJ, Upton JH 3rd, Rodarte L, Romero TA, et al. (2010). A 3' terminal stem-loop structure in Nodamura virus RNA2 forms an essential *cis*-acting signal for RNA replication. *Virus Res.* 150: 12-21.
- Sean P, Nguyen JHC and Semler BL (2009). Altered interactions between stem-loop IV within the 5' noncoding region of coxsackievirus RNA and poly(rC) binding protein 2: effects on IRES-mediated translation and viral infectivity. *Virology* 389: 45-58.
- Ueno M, Kodama EN, Shimura K, Sakurai Y, et al. (2009). Synonymous mutations in stem-loop III of Rev responsive elements enhance HIV-1 replication impaired by primary mutations for resistance to enfuvirtide. *Antiviral Res.* 82: 67-72.

- Wang XF, Wang X, Liu J, Feng J, et al. (2009a). Alu tandem sequences inhibit GFP gene expression by triggering chromatin wrapping. *Genes Genom.* 31: 209-215.
- Wang XF, Jin X, Wang X, Liu J, et al. (2009b). Effects of L1-ORF2 fragments on green fluorescent protein gene expression. *Genet. Mol. Biol.* 32: 688-696.
- Xu ZL, Mizuguchi H, Ishii-Watabe A, Uchida E, et al. (2001). Optimization of transcriptional regulatory elements for constructing plasmid vectors. *Gene* 272: 149-156.
- Yin K, Wang X, Ma H, Xie Y, et al. (2010). Impact of copy number of distinct SV40PolyA segments on expression of a GFP reporter gene. *Sci. China Life Sci.* 53: 606-612.
- Yu L and Markoff L (2005). The topology of bulges in the long stem of the flavivirus 3'stem-loop is a major determinant of RNA replication competence. *J. Virol.* 79: 2309-2324.