Available online at www.sciencedirect.com
SCIENCE DIRECT
Bioorganic & Medicinal Chemistry Letters
Bioorganic & Medicinal Chemistry Letters 16 (2006) 1410–1413

Synthesis of linked triple helical DNAs possessing high affinity to triple helical DNA binding protein

Aya Shibata,[a] Yoshihito Ueno,[a] Koichiro Shinbo,[a] Masayuki Nakanishi,[a] Akira Matsuda[b] and Yukio Kitade[a,*]

[a]*Department of Biomolecular Science, Faculty of Engineering, Gifu University, Yanagido, Gifu 501-1193, Japan*
[b]*Graduate School of Pharmaceutical Sciences, Hokkaido University, Kita-12, Nishi-6, Kita-ku, Sapporo 060-0812, Japan*

Received 30 September 2005; revised 8 November 2005; accepted 10 November 2005
Available online 5 December 2005

Abstract—The synthesis of triplexes possessing an anthraquinonyl group and composed of branched oligonucleotides (ONs) is described. Binding ability of a triplex-binding protein (MBP-LOR3^{ARF}) to the triplexes was evaluated by an electrophoretic mobility shift assay (EMSA). It was found that the triplex, which has an anthraquinonecarbonyl group at the 5′-end of the third strand and is connected with the pentaerythritol linker, has greater affinity to the protein than an unmodified triplex.

Triple helix (triplex)-forming oligonucleotides (ONs) have attracted a great deal of attention because of their ability to specifically bind double-stranded DNA and their potential use in gene therapy.[1,2] Depending on the orientation of the third strand, two major classes of triplexes are identified.[3] When the third strand consists mainly of pyrimidines, Hoogsteen-type base triplets (T·dA:T and dC^{+}·dG:dC) are formed in which the third strand is parallel to the purine strand of the target duplex. When the third strand is predominantly purines, reverse Hoogsteen-type base triplets (dG·dG:dC, dA·dA:T, and T·dA:T) are formed in which the third strand is antiparallel to the purine strand of the target duplex.

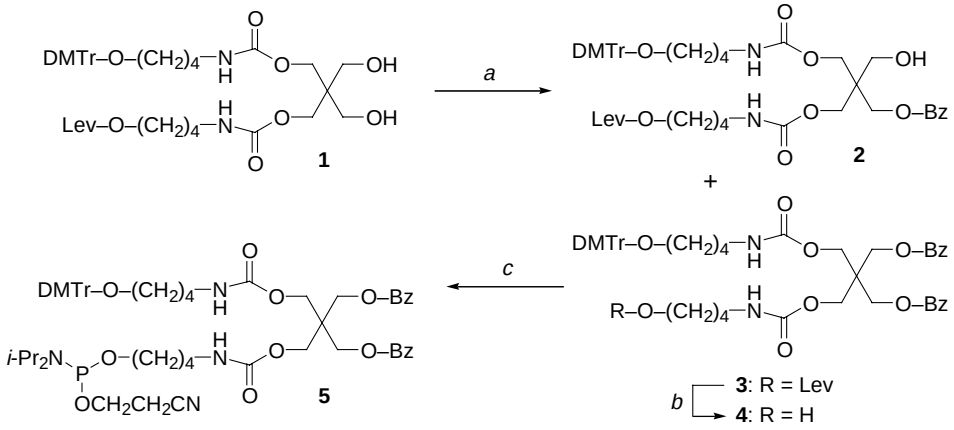
Recently, much attention is being given to methodology targeting single-stranded DNA or RNA through the formation of triplexes using circular or branched ONs.[4–10] For instance, Volkmann et al. showed that in cell culture experiments a branched ON against a highly conserved polypurine tract efficiently inhibited HIV-1 replication, leading to a block of p24 synthesis and inhibition of syncytia formation.[9] Recently, we have also demonstrated the synthesis of the branched ONs that were linked with the pentaerythritol linker.[10] We found that the branched ONs linked with the pentaerythritol linker significantly stabilized the triplexes with single-stranded DNA or RNA.

On the other hand, so far, proteins, that specifically bind triple helical DNAs, have been found in several kinds of species.[11–21] They are referred to as triplex-binding proteins (TBPs). Although the biological roles of the TBPs are not fully understood, they are presumed to play important roles in cells by binding the triple helical DNAs. From these findings, we envisioned that the TBPs would be good stabilizers of triplexes in the antisense strategy utilizing the branched ONs if they specifically bind triplexes between the branched ONs and the target DNAs or RNAs, and thermally stabilize the triple helical structures. In this paper, we address the possibility of the TBPs as stabilizers of triplexes in the antisense strategy utilizing the branched ONs.

In order to synthesize branched ONs connected with the pentaerythritol linker, a phosphoramidite of a pentaerythritol derivative was synthesized according to Scheme 1. Hydroxyl functions of a properly protected pentaerythritol **1**, which was synthesized by the reported method,[10,22–25] were protected with a Bz group to give a di-Bz derivative **3** in 86% yield along with the formation of a mono-Bz derivative **2**. After deprotecting the Lev group of **3**, **4** was phosphitylated by the standard procedure[26] to give the corresponding phosphoramidite **5** in 94% yield. To incorporate an intercalator at the 5′-end of a third strand, a phosphoramidite of an anthraquinone derivative was synthesized

Keywords: Triple helical DNA; Triplex binding protein; Pentaerythritol linker; Anthraquinonyl group; Electrophoretic mobility shift assay.
* Corresponding author. Tel./fax: +81 58 293 2640; e-mail: kitade@biomol.gifu-u.ac.jp

0960-894X/$ - see front matter © 2005 Elsevier Ltd. All rights reserved.
doi:10.1016/j.bmcl.2005.11.036



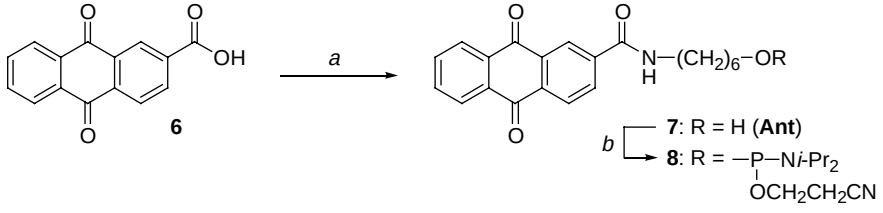
Scheme 1. Reagents and conditions: (a) BzCl, pyridine, rt, 86%; (b) $NH_2NH_2{\cdot}H_2O$, CH_3CO_2H, pyridine, rt, 81%; (c) $ClP(OC_2H_4CN)(N\text{-}i\text{-}Pr_2)$, $EtN\text{-}i\text{-}Pr_2$, CH_2Cl_2, rt, 94%.

according to Scheme 2. The anthraquinone derivative **7**, which was obtained by a reaction of anthraquinone-2-carboxylic acid (**6**) and 6-aminohexanol, was phosphitylated by the standard procedure[26] to give the corresponding phosphoramidite **8** in 85% yield. ONs were synthesized using the phosphoramidite method with a DNA/RNA synthesizer. Sequences of ONs used in this study are listed in Figure 1 (The clusters of 2′-deoxyguanosine are abbreviated to G4, G3 or G2 in the triplexes). The ONs were designed to form antiparallel-type triplexes. The structure of the ONs was analyzed by MALDI-TOF/MS, and the observed molecular weight supported its structure.

We chose the MBP-LOR3^{ARF} protein, which was found in South-Western screening of a human keratinocyte expression library by Ciotti et al.,[27] as a TBP for this study. First, we evaluated the binding ability of the MBP-LOR3^{ARF} protein to the non-linked triplexes by an electrophoretic mobility shift assay (EMSA). The triplexes labeled with ^{32}P at the 5′-ends of the first strands were incubated with various concentrations of the MBP-LOR3^{ARF} protein, and the mixtures were analyzed by a polyacrylamide gel electrophoresis (PAGE) under non-denaturing conditions.[28] As shown in Figure 2, the PAGE exhibited two slower mobility bands, which would correspond to the triplex-protein complexes. The result is consistent with the previous report.[20] It was suggested that 5′-GGG-3′ is the minimum sequence recognized by an individual MBP-LOR3^{ARF} molecule. Thus, the faster mobility band referred to as C1 is supposed to correspond to the complex composed of the triplex and single or plural proteins, whereas the slower mobility band referred to as C2 is presumed to correspond to the complex between the triplex and plural proteins. The densities of radioactive bands on the gel were determined with a bio-imaging analyzer. The

Abbreviation of triplex	ONs	Sequence
G3-G4-G4-G3	ON1	5′-TGGGTGGGGTGGGGTGGGT-3′
	ON2	3′-AGGGAGGGGAGGGGAGGGA-5′
	ON3	5′-TCCCTCCCCTCCCCTCCCT-3′
G3-G3-G3-G3	ON4	5′-TGGGTGGGTTTGGGTGGGT-3′
	ON5	3′-AGGGAGGGAAAGGGAGGGA-5′
	ON6	5′-TCCCTCCCTTTCCCTCCCT-3′
G2-G4-G4-G2	ON7	5′-TTGGTGGGGTGGGGTGGTT-3′
	ON8	3′-AAGGAGGGGAGGGGAGGAA-5′
	ON9	5′-TTCCTCCCCTCCCCTCCTT-3′
G2-G4-G2-G4	ON10	5′-TTGGTGGGGTTGGTGGGGT-3′
	ON11	3′-AAGGAGGGGAAGGAGGGGA-5′
	ON12	5′-TTCCTCCCCTTCCTCCCCT-3′
Ant-G3-G4-G4-G3*	ON13	5′-**Ant**-TGGGTGGGGTGGGGTGGGT-3′
	ON14	3′-AAAGAGGGAGGGGAGGGGAGGGA-5′
	ON15	5′-TTTCTCCCTCCCCTCCCCTCCCT-3′
linked-G3-G4-G4-G3	ON16	5′-TGGGTGGGGTGGGGTGGGT-OH 3′-AGGGAGGGGAGGGGAGGGA-OH
	ON3	5′-TCCCTCCCCTCCCCTCCCT-3′
Ant-linked-G3-G4-G4-G3*	ON17	5′-**Ant**-TGGGTGGGGTGGGGTGGGT-OH 3′-AAAGAGGGAGGGGAGGGGAGGGA-OH
	ON15	5′-TTTCTCCCTCCCCTCCCCTCCCT-3′
mis G3-G4-G4-G3	ON1	5′-TGGGTGGGGTGGGGTGGGT-3′
	ON2	3′-AGGGAGGGGAGGGGAGGGA-5′
	ON18	5′-TCCCTCCCCTC**G**CCTCCCT-3′
mis linked-G3-G4-G4-G3	ON16	5′-TGGGTGGGGTGGGGTGGGT-OH 3′-AGGGAGGGGAGGGGAGGGA-OH
	ON18	5′-TCCCTCCCCTC**G**CCTCCCT-3′

Figure 1. Sequences of the oligonucleotides used in this study.



Scheme 2. Reagents and conditions: (a) $NH_2(CH_2)_6NH_2$, 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide hydrochloride, *N*-methylimidazole, DMF, rt, 25%; (b) $ClP(OC_2H_4CN)(N\text{-}i\text{-}Pr_2)$, $EtN\text{-}i\text{-}Pr_2$, CH_2Cl_2, rt, 85%.

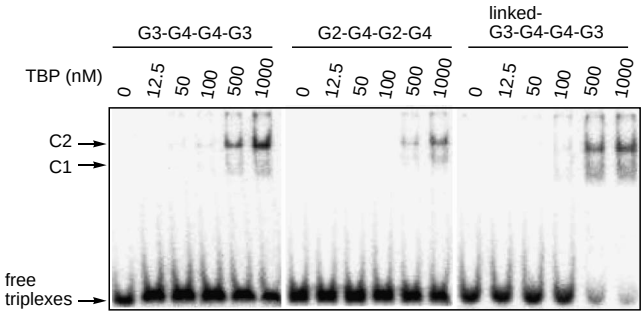


Figure 2. Quantitative EMSA to detect binding of TBP to the triplexes, **G3–G4–G4–G3**, **G2–G4–G2–G4**, and **linked-G3–G4–G4–G3**. Triplex concentration: 50 nM. TBP concentrations are indicated.

results are shown in Figure 3 by a bar graph. It was found that the protein has a slightly greater affinity to the triplexes composed of the continuous 5′-GGGG-3′ sequences than others.

Next, we appraised the binding ability of the MBP-LOR3^{ARF} protein to the triplexes modified with the anthraquinone derivative and comprised of the branched ON **16** or ON **17**. The binding ratios of the protein to the triplexes are shown in Figure 4. It was revealed that the linked triplex has a higher affinity to the protein than the non-linked triplexes at every protein concentration. Furthermore, it turned out that the incorporation of the anthraquinonyl group also increases affinity of the triplexes to the protein. It was reported that thermal stabilities of triplexes were enhanced by connecting the third strands and the first or second strands with proper linkers and introducing intercalators into the third strands.[4,5,22] Thus, the increased affinity of the protein to the triplexes would be due to the enhancement of thermal stability of the triplexes.

To examine whether the binding of the protein is specific for a triplex structure or not, we performed the EMSA in the presence of competitors. The triplex, **linked-G3-G4-G4-G3**, labeled with ^{32}P at the 5′-end of the first strand was incubated with the MBP-LOR3^{ARF} protein at room temperature for 60 min. The mixture was further

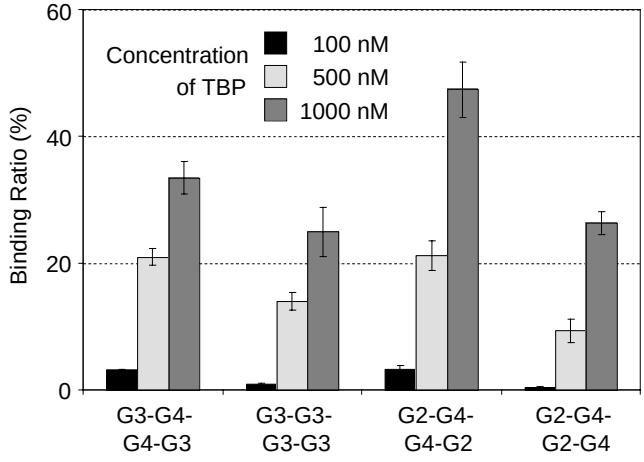


Figure 3. Binding ratio of TBP to the triplexes, **G3–G4–G4–G3**, **G3–G3–G3–G3**, **G2–G4–G4–G2**, and **G2–G4–G2–G4**.

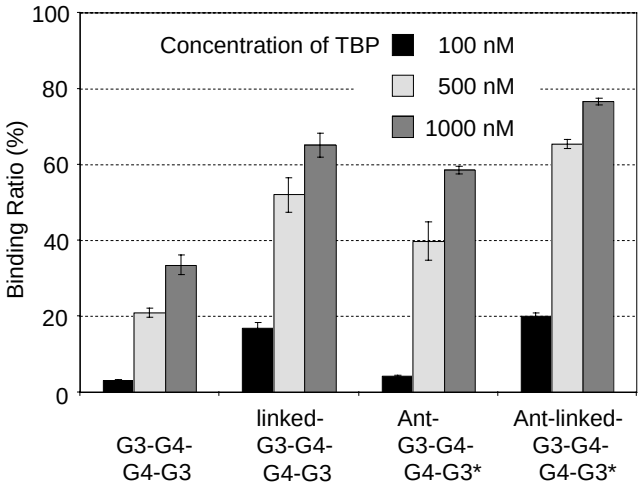


Figure 4. Binding ratio of TBP to the triplexes, **G3–G4–G4–G3**, **Ant-G3–G4–G4–G3***, **linked-G3-G4–G4–G3**, and **Ant-linked-G3–G4–G4–G3***.

incubated in the presence of the competitors for 30 min and was analyzed by the PAGE under non-denaturing conditions. The results are shown in Figure 5. Binding ratio of the protein to the labeled triplex was decreased in the presence of the triplexes, **G3-G4-G4-G3** and **linked-G3-G4-G4-G3**, whereas other competitors had almost no influence on the binding of the protein to the labeled triplex. Thus, it was revealed that the binding of the protein is specific for the triplex structure.

Finally, we investigated the binding ability of the MBP-LOR3^{ARF} protein to the triplexes containing one base mismatch in the first strands. As shown in Figure 6, it was found that the incorporations of the mismatch base pair significantly reduce the affinity of the protein to the triplexes both in the linked and non-linked triplexes.

In conclusion, we have demonstrated a synthesis of triplexes composed of the branched ONs and modified

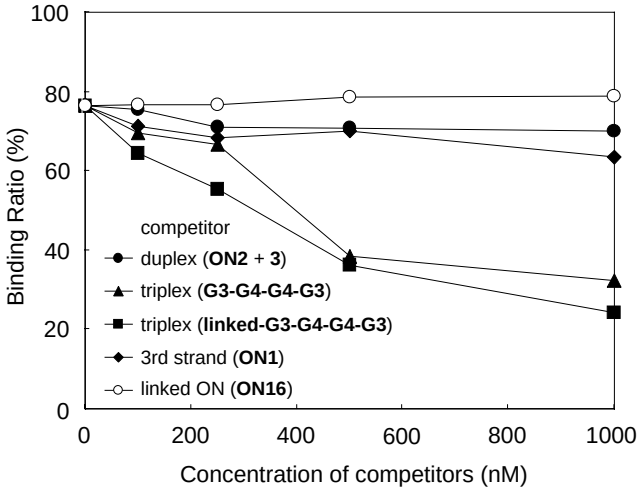


Figure 5. Binding ratio of TBP to the ^{32}P-labeled **linked-G3–G4–G4–G3** in the presence of the competitors. Concentration of ^{32}P-labeled **linked-G3–G4–G4–G3**: 50 nM. Concentration of the TBP: 1000 nM. Concentrations of the competitors are indicated.

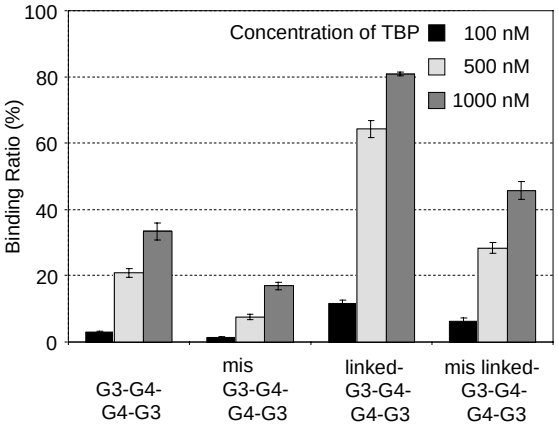


Figure 6. Binding ratio of TBP to the triplexes, **G3–G4–G4–G3**, **mis-G3–G4–G4–G3**, **linked-G3–G4–G4–G3**, and **mis-linked-G3–G4–G4–G3**.

with the anthraquinonyl group. The binding ability of the MBP-LOR3^{ARF} protein to the triplexes was evaluated by the EMSA. It was found that the triplex, which has an anthraquinonecarbonyl group at the 5′-end of the third strand and is connected with the pentaerythritol linker, has greater affinity to the protein than an unmodified triplex. Potaman and Sinden reported that peptides composed of basic amino acids such as Lys and Arg thermally stabilize triplexes.[29,30] The MBP-LOR3^{ARF} protein also has clusters comprised of basic amino acids at its C-terminus region. Thus, the MBP-LOR3^{ARF} protein will be a good stabilizer of triplexes in antisense strategy utilizing the branched ONs.

Acknowledgment

This research was supported in part by a Grant-in-Aid for Scientific Research (C) (KAKENHI 16590082) from Japan Society for the Promotion of Science (JSPS).

References and notes

1. Praseuth, D.; Guieysse, A. L.; Hélèn, C. *Biochim. Biophys. Acta* **1999**, *1489*, 181.
2. Giovannangeli, C.; Hélèn, C. *Nat. Biotechnol.* **2000**, *18*, 1245.
3. Thuong, N. T.; Hélène, C. *Angew. Chem., Int. Ed. Engl.* **1993**, *32*, 666.
4. Giovannangeli, C.; Montenay-Garestier, T.; Rougée, C. M.; Thuong, N. T.; Hélène, C. *J. Am. Chem.Soc.* **1991**, *113*, 7775.
5. Giovannangeli, C.; Thuong, N. T.; Hélène, C. *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 10013.
6. Kool, E. T. *Chem. Rev.* **1997**, *97*, 1473.
7. Wang, S.; Kool, E. T. *Nucleic Acids Res.* **1995**, *23*, 1157.
8. Kandimalla, E. R.; Agrawal, S.; Venkataraman, G.; Sasisekharan, V. *J. Am. Chem. Soc.* **1995**, *117*, 6416.
9. Volkmann, S.; Jendis, J.; Frauendorf, A.; Moelling, K. *Nucleic Acids Res.* **1995**, *23*, 1204.
10. Ueno, Y.; Shibata, A.; Matsuda, A.; Kitade, Y. *Bioconjugate Chem.* **2003**, *14*, 684.
11. Kiyama, R.; Camerini-Otero, R. D. *Proc. Natl. Acad. Sci. U.S.A.* **1991**, *88*, 10450.
12. Frantz, J. D.; Gilbert, W. *J. Biol. Chem.* **1995**, *270*, 9413.
13. Guieysse, A. L.; Praseuth, D.; Hélèn, C. *J. Mol. Biol.* **1997**, *267*, 289.
14. Jiménez-Garacía, E.; Vaquero, A.; Espinás, M. L.; Soliva, R.; Orozco, M.; Bernués, J.; Azorín, F. *J. Biol. Chem.* **1998**, *273*, 24640.
15. Musso, M.; Nelson, L. D.; Van Dyke, M. W. *Biochemistry* **1998**, *37*, 3086.
16. Musso, M.; Bianchi-Scarrà, G.; Van Dyke, M. W. *Nucleic Acids Res.* **2000**, *28*, 4090.
17. Nelson, L. D.; Musso, M.; Van Dyke, M. W. *J. Biol. Chem.* **2000**, *275*, 5573.
18. Guillonneau, F.; Guieysse, A. L.; Le Caer, J. P.; Rossier, J.; Praseuth, D. *Nucleic Acids Res.* **2001**, *29*, 2427.
19. Rao, J. E.; Craig, N. C. *J. Mol. Biol.* **2001**, *307*, 1161.
20. Ciotti, P.; Van Dyke, M. W.; Bianchi-Scarrà, G.; Musso, M. *Eur. J. Biochem.* **2001**, *268*, 225.
21. Li, G.; Tolstonog, G. V.; Traub, P. *DNA Cell Biol.* **2002**, *21*, 163.
22. Ueno, Y.; Takeba, M.; Mikawa, M.; Matsuda, A. *J. Org. Chem.* **1999**, *64*, 1211.
23. Ueno, Y.; Mikawa, M.; Hoshika, S.; Matsuda, A. *Bioconjugate Chem.* **2001**, *12*, 635.
24. Hoshika, S.; Ueno, Y.; Matsuda, A. *Bioconjugate Chem.* **2003**, *14*, 607.
25. Kim, S. J.; Bang, E.-K.; Kim, B. H. *Synlett* **2003**, 1838.
26. Sinha, N. D.; Biernat, J.; Köster, H. *Tetrahedron Lett.* **1983**, *24*, 5843.
27. The MBP-LOR3^{ARF} protein was prepared as follows: The cDNA encoding LOR3^{ARE} fragment was PCR-amplified from human keratinocyte cDNA library (Clontech) using a primer set, 5′-GGGGATCCGAGGGGGGTCGTCCG GCG-3′ and 5′-GCGAATTCCATGAGAGCTCTAAG CCCATC-3′, as described by Ciotti et al.[20] The PCR product was cloned to pET-21a expression vector at the *Bam*HI and *Eco*RI restriction sites (pET-LOR3^{ARF}). A 0.3 kb-DNA fragment was obtained from pET- LOR3^{ARF} by digesting with *Bam*HI and *Hind*III, and was subcloned to pEM vector (pEM- LOR3^{ARF}). pEM vector was constructed by inserting MBP cDNA in pET-22b. MBP cDNA was obtained by PCR with a primer set, 5′-ATAGCATATGAAAATCGAAGAAGGT-3′ and 5′-CGCTTCTGCGTTCTGATTTA-3′, using pMAL-c2 vector (New England Biolabs) as a template. LOR3^{ARF} was expressed as MBP N-terminal and 6× His C-terminal fusion protein in *Escherichia coli* strain JM109 (DE3). Bacteria were grown in LB containing 50 μg/ml ampicillin at 30 °C. The expression was induced by adding 0.5 mM IPTG to the culture at 25 °C. The recombinant protein was purified from *E. coli* using TALON™ Metal Affinity Resins followed by Amylose Resin™.
28. Electrophoretic mobility shift assays: Triplex (50 nM) labeled at the 5′-end of the 1st strand by ^{32}P was incubated in the presence of various amounts of the MBP-LOR3^{ARF} protein in a 15 mM Tris–HCl buffer (pH 7.0) containing 25 mM NaCl, 5 mM $MgCl_2$, and 10% glycerol (total 20 μL) at room temperature for 1 h. The mixture was analyzed by electrophoresis on 5% nondenaturing polyacrylamide gel (19:1 acrylamide/bisacrylamide) using a TB buffer (22 mM Tris borate, pH 8, 0.5 mM EDTA) at 80–100 mV for 2 h at 4 °C. Density of radioactivity of the gel was visualized by a Bio-imaging Analyzer (Bas 2500, Fuji, Co. Ltd.).
29. Potaman, V. N.; Sinden, R. R. *Biochemistry* **1995**, *34*, 14885.
30. Potaman, V. N.; Sinden, R. R. *Biochemistry* **1998**, *37*, 12952.