

G-Rich Sequence-Specific Recognition and Scission of Human Genome by PNA/DNA Hybrid G-Quadruplex Formation**

Takumi Ishizuka, Jie Yang, Makoto Komiyama,* and Yan Xu*

Recognition and scission of sequences within duplex DNA are widely used for gene manipulation.^[1] To date, naturally occurring restriction enzymes have been used. However, the recognition site sizes (4–6 bp) of most natural restriction enzymes limit their usefulness for many applications. For example, restriction enzymes that recognize 6 bp long sequences cut human DNA (which contains about 3 billion bp) at more than 10^5 sites on average. To increase sequence specificity, numerous attempts have been made to cleave duplex DNA at specific sites.^[2] Protein-based artificial nucleases were engineered to target the desired DNA sequences.^[3] Recently, chemistry-based approaches have been effective for sequence-specific recognition of DNA through base-pairing.^[4] For example, several research groups, including our own, have shown that pseudo-complementary PNA (pcPNA) is effective in recognizing complementary AT-rich sequences using the modified bases 2,6-diaminopurine (D) and 2-thiouracil (U_s).^[5] Although these modified PNAs have recently been developed to target AT-rich sequences by a double-duplex invasion strategy, this strategy is known to be difficult to target guanine-rich (G-rich) sequences because the modified bases D and U_s cannot be used owing to the lack of AT bases in the G-rich sequence.^[5a] Therefore, an effective

chemical method to overcome this limitation for efficient cleavage G-rich sequences is desired.

G-rich sequences are located in telomeres as repeated units as well as elsewhere in the genome.^[6] G-rich sequences have been proposed to have important biological roles through the formation of G-quadruplex structures. Bioinformatic analysis has identified 375 000 G-rich candidate sequences within the human genome capable of folding into G-quadruplexes.^[7] For example, the promoter in the G-rich sequence of the *c-myc* oncogene forms a G-quadruplex and controls 85–90 % of transcriptional activation.^[8] Developing various approaches to target such G-rich regions is of great interest because of their potential role in discovering anti-cancer therapeutic agents and other versatile purposes.^[9] Many studies have focused on using molecules to selectively bind G-quadruplexes.^[10]

Herein, we present a novel approach for targeting G-rich double-stranded DNA (dsDNA). The method was successfully applied to sequence-specific DNA-breaking in the target sequence. As shown in Figure 1 a, G-rich PNA can bind to a homologous DNA strand, such as the *c-myc* oncogenic promoter sequence, by forming a hybrid PNA/DNA G-quadruplex. The G-rich PNA is automatically complementary to the other DNA strand by Watson–Crick duplex formation. Thus, a single PNA can bind to both DNA strands at the same site. The distinct advantage of the method is that with the use of G-quadruplex formation between PNA and DNA, only one unmodified PNA has the potential to recognize both strands of G-rich dsDNA. We further apply this method to sequence-specific breaking in the target sequence of human genome (Figure 1 b). We show that only one target site in the whole genome of human beings (one site in the chromosome 8) was selectively hydrolyzed by PNA/DNA hybrid G-quadruplex formation. The chemical strategy overcomes the limitation of natural restriction enzymes and may serve as a new technology for gene manipulation.

As shown in Figure 2 a, we synthesized a series of G-rich PNA strands that were designed to bind to the target dsDNA (Supporting Information, Table S1, Figure S1). We first assessed their strand invasion ability in PNA/DNA complex formation using a gel shift assay.^[10d,11] A 99-base pair (bp) PCR fragment, DNA1 (Figure 2 a), containing the half-length G-rich sequence of the *Myc19* region was incubated with PNA1 at 37 °C and pH 7.0 in the presence of 10 mM NaCl. We observed that strand invasion in the complex formation was effective; the intensity of the new band gradually increased and the starting material (DNA1) decreased with increasing PNA1 concentration (lanes 1–4 in Figure 2 b). Furthermore, we found that PNA1 effectively binds to DNA2 (108 bp, Figure 2 a) containing the full-length G-rich sequence of the

[*] Dr. T. Ishizuka,^[+] Prof. Dr. M. Komiyama^[++]
Research Center for Advanced Science and Technology
The University of Tokyo
4-6-1 Komaba, Meguro-ku, Tokyo 153-8904 (Japan)
E-mail: komiyama@ara.tsukuba.ac.jp
J. Yang, Prof. Dr. Y. Xu
Division of Chemistry, Department of Medical Sciences
Faculty of Medicine, University of Miyazaki
5200 Kihara, Kiyotake, Miyazaki 889-1692 (Japan)
E-mail: xuyan@fc.miyazaki-u.ac.jp
Homepage: http://www.miyazaki-med.ac.jp/MMCHEMA/index_e_link_e.html

[+] Current address: RIKEN Systems and Structural Biology Center (SSBC), 1-7-22 Suehiro-cho, Tsurumi-ku
Yokohama, Kanagawa 230-0045 (Japan)

[++] Current address: Life Science Center of Tsukuba Advanced Research Alliance University of Tsukuba
1-1-1 Tennoudai, Tsukuba, Ibaraki 305-8577 (Japan)

[**] This work was supported in part by a Grant-in-Aid for Specially Promoted Scientific Research (18001001) and a Grant-in-Aid for Scientific Research (B) (22350071) from the Ministry of Education, Science, Sports, Culture and Technology (Japan), and by the Research Fellowships of the Japan Society for the Promotion of Science for Young Scientists. Support from the Takeda Science Foundation is also acknowledged.

Supporting information for this article is available on the WWW under <http://dx.doi.org/10.1002/anie.201201176>.

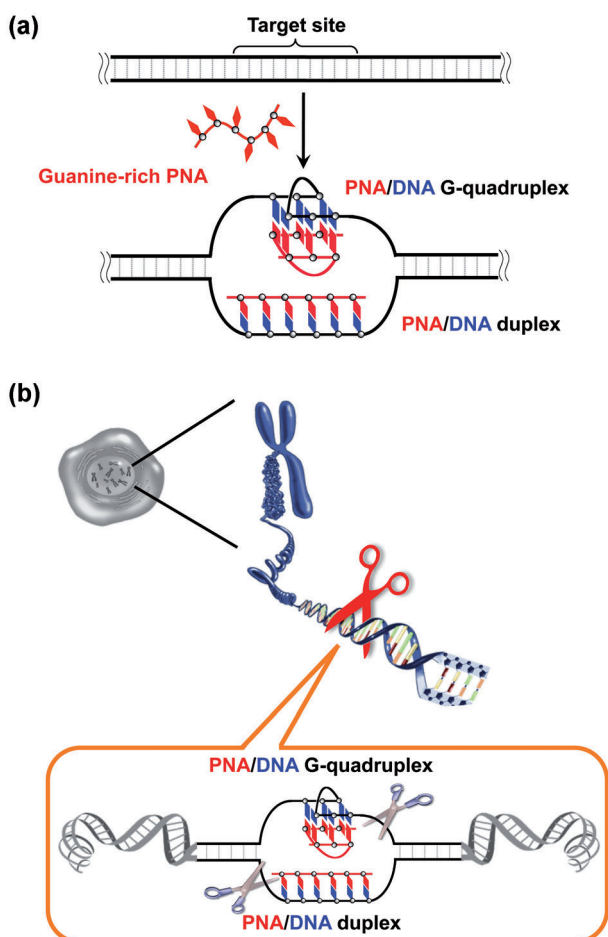


Figure 1. G-rich sequence-specific recognition and scission by PNA/DNA hybrid G-quadruplex formation. a) Representation showing that PNA/DNA hybrid G-quadruplex formation between G-rich PNA and homologous DNA of the upper target strand promotes a single PNA to effectively target the G-rich region of the *c-myc* promoter of human oncogene; a duplex is formed at the lower strand by the same complementary PNA strand. b) The present G-rich sequence-specific scission of the whole human genome at *c-myc* gene promoter in the chromosome.

Myc19 region, which has been known to form intramolecular G-quadruplexes (lanes 6–8 in Figure 2b).

To further confirm that single PNA binds to both strands of G-rich dsDNA, more rigorous investigations were performed to examine the binding ability using mutated DNA substrates and PNAs. We prepared a mutated DNA3 substrate that did not form a G-quadruplex; in this mutated substrate, two dG residues in the G-rich sequence of the DNA1 substrate were substituted with dAs (Figure 2a). When PNA1m that was designed to be complementary to the lower strand of the DNA3 target by duplex formation was used in a parallel experiment, no shifted band was observed (lane 3 in Figure 2c) compared to the DNA1 substrate and PNA1 (lane 2 in Figure 2c) containing a homologous G-rich sequence at the upper strand and a complementary sequence at the lower strand. A control experiment was performed to further confirm the formation of G-quadruplex between PNA and DNA. A mutated DNA5 substrate (Supporting Informa-

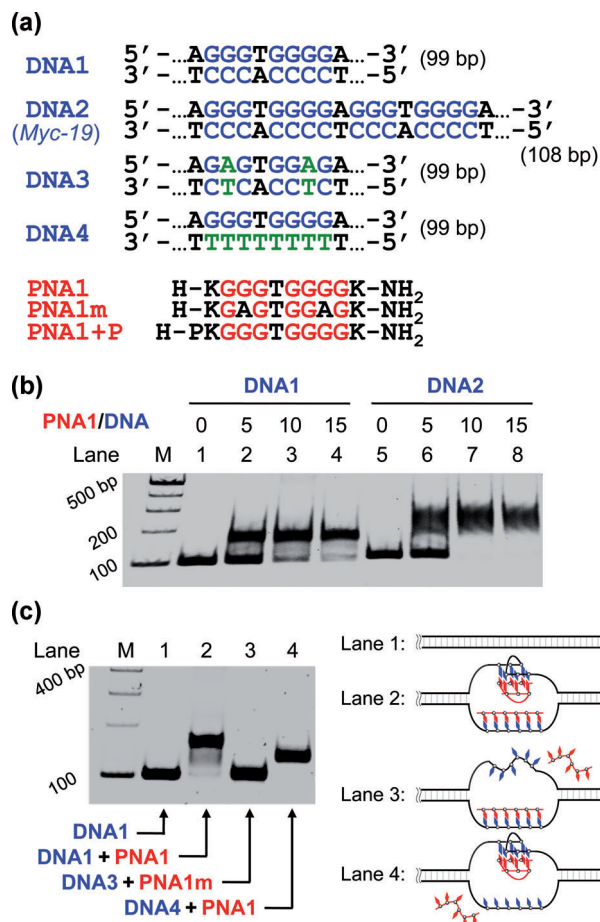


Figure 2. PNA/DNA G-quadruplex promoting the invasion of a single PNA to duplex DNA. a) All DNA sequences used in this study. b), c) Gel shift assay for site-selective recognition of G-rich PNA. Conditions for (b): [DNA1] or [DNA2] = 50 nM, [PNA1] = 0–1 μM, [HEPES buffer] = 5 mM (pH 7.0), [NaCl] = 10 mM, 37°C. lane M: marker, lane 1: DNA1 only, lanes 2–4: DNA1 + PNA1, lane 5: DNA2 only, lanes 6–8: DNA2 + PNA1. PNA1/DNA indicates the molar ratio of PNA1 to DNA1 or DNA2. For (c): Lane M: marker, lane 1: control DNA1, lane 2: DNA1 + PNA1, lane 3: mutated DNA3 substrate at the upper strand + PNA1m (unable to form G-quadruplex), lane 4: mutated DNA4 at the lower strand + PNA1. Conditions are the same as described in (b).

tion, Figure S2) having the same sequence of lower strand DNA1 that can form a duplex with PNA1 at the lower strands but not G-quadruplex formation at upper was used, we found that the shifted band of DNA5 and PNA1 is lower than that of the DNA1 substrate and PNA1 (Supporting Information, Figure S2, lane 3). A concentration-dependence experiment showed that the PNA/DNA hybrid G-quadruplex is effectively formed at low NaCl concentrations, consistent with previous studies that G-quadruplex is stable at low salt concentrations (Supporting Information, Figure S3). These suggest that formation of a G-quadruplex between PNA and target DNA is a key step in the binding interaction. Moreover, when a mutated DNA4 substrate (all dT substitution sequence at the lower strand of the target site, Figure 2a) that did not form a duplex was used, we noticed that the shifted band had half the mobility (lane 4 in Figure 2c) compared to

that of the DNA1 substrate and PNA1 (lane 2 in Figure 2c). This suggests that formation of a duplex between PNA and DNA is needed for the stabilized invasion complexes. Taken together, these observations imply that the recognition of both the upper and lower strands is important for the formation of invasion complexes, indicating that the targeting approach to G-rich sequence is site-specificity. We next evaluated the kinetics of the complex formation by G-rich PNA invasion. We found that the efficiency of the present invasion is dependent on temperature. A time-course study demonstrated that formation of this complex followed pseudo first-order kinetics (Supporting Information, Figure S4).

To obtain further evidence for intermolecular PNA/DNA G-quadruplex formation, we performed a dimethyl sulfate (DMS) footprinting assay. DMS footprinting is used to identify G residues involved in the formation of the G-quadruplex. The N7 position of each G involved in the formation of a G-quadruplex is protected against methylation by DMS through Hoogsteen bonding.^[12] The cleavage pattern of G residues is used to probe G-quadruplex formation. In the absence of PNA1, a sequence containing the target (AGGGTGGGGA) was cleaved randomly at every G residue (Supporting Information, Figure S5, lanes 1–3; sequences of ODNs are shown in Table S2), suggesting an unstructured form. However, in the presence of PNA1, G residues were either partially or fully protected from piperidine cleavage (SI, Figure S5, lanes 4–6). The DMS footprinting pattern suggests that PNA1 and DNA substrate can form an intermolecular PNA/DNA G-quadruplex. We performed the additional experiments to further clarify the mechanism of G-quadruplex protection. PNAs that did not form a G-quadruplex and only form duplex with substrates did not protect against DMS methylation, suggesting that the protected methylation results from G-quadruplex formation (Supporting Information, Figure S6, S7).

We employed circular dichroism (CD) spectroscopy to investigate the thermal stability (T_m) of the key components of PNA/DNA G-quadruplex at the target site (Supporting Information, Figure S8).^[13] According to CD melting experiments, T_m of the PNA/DNA G-quadruplex (61.0 °C) was notably higher than that of the DNA/DNA duplex (40.3 °C). Thus, PNA/DNA G-quadruplex formation at the homologous target DNA shows high propensity toward invasion in this system. This is consistent with the result that PNA1 can overcome the interrupted duplex DNA structure on the target to form the PNA/DNA G-quadruplex.

Having established this efficient invasion of the PNAs to duplex DNA, we were encouraged to apply this method to sequence-specific double-strand breaking in the target sequence of plasmid DNA (Figure 3a). Our strategy involves breaking the target DNA sequence by forming G-quadruplex and duplex between PNA and target DNA.^[4a] To promote efficient scission, an L-phosphoserine (P) as a monophosphate group was attached to the N-termini of G-rich PNA (PNA1 + P) (Figure 3a; Supporting Information, Figure S9). Because phosphonate-type ligands exhibit high affinity toward lanthanide ions, PNAs bearing this type of terminal group efficiently recruited the catalytic Ce^{IV} /EDTA species to the target site, resulting in efficient cleavage.^[14] Using the site-directed

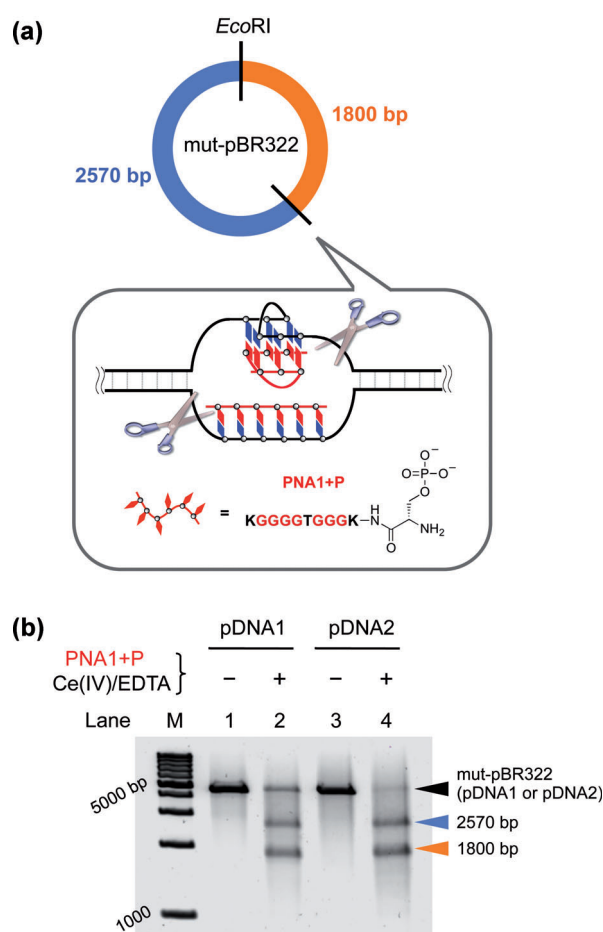


Figure 3. G-rich sequence-specific breaking of double strand in plasmid DNA by G-rich PNA and Ce^{IV} /EDTA complex. a) Vector map of mut-pBR322 (pDNA1 or pDNA2) containing the target sequences (DNA1 and DNA2) for scission as shown in Figure 2a. Lengths of fragments (1800 and 2570 bp) obtained by consecutive scissions by G-rich PNA (at target sites) and EcoRI are shown. b) Gel electrophoresis patterns. Arrows indicate the two fragments (1800 and 2570 bp). lane M: 1 kbp DNA ladder, lane 1: EcoRI digest for pDNA1, lane 2: after sequence-specific scission + EcoRI digest for pDNA1, lane 3: EcoRI digest for pDNA2, lane 4: after sequence-specific scission + EcoRI digest for pDNA2. Reaction conditions: target [pDNA1] or [pDNA2] = 4 nM, $[\text{Ce}^{\text{IV}}/\text{EDTA}] = 200 \mu\text{M}$, $[\text{PNA1} + \text{P}] = 200 \text{ nM}$, $[\text{HEPES}] = 5 \text{ mM}$, and $[\text{NaCl}] = 10 \text{ mM}$ at pH 7.0 and 50 °C for 16 h.

mutagenesis method, we prepared two types of plasmid DNAs pDNA1 and pDNA2 (Figure 3a) containing the target sites DNA1 and DNA2, respectively. Precise construction of site-directed mutagenesis pBR322 vectors containing the target sequence and EcoRI site were directly confirmed by sequencing the extracted plasmid (Supporting Information, Figure S10). The invasion complex of pDNA1 and pDNA2 substrate with PNA1 + P was treated with Ce^{IV} /EDTA at pH 7.0 and 50 °C for 16 h and then digested with EcoRI . The results of agarose gel electrophoresis are presented in Figure 3b (lanes 2 and 4). Only two scission bands were observed, at around 1.8 and 2.5 kbp, which is exactly as expected from site-selective scission in this case (1800 and 2570 bp fragments should be formed). The pBR322 vector

that does not contain the target sequences (DNA1 and DNA2) was used as a control experiment. We noticed that in the presence of PNA1, the pBR322 vector was not cleaved (Supporting Information, Figure S11). A further experiment on cleavage efficiency as a function of Ce^{IV} /EDTA concentration was performed in the concentration range from 0 to 200 μM (Supporting Information, Figure S12). The cleavage efficiency increased significantly with increasing Ce^{IV} /EDTA concentration. The yield of scission fragments amounted to 87 % at a concentration of 200 μM (Supporting Information, Figure S13). Higher efficiency at lower concentrations will be preferable in such PNAs for cellular use.

Encouraged by these data, we next tested whether this method is effective on sequence-specific scission of the whole human genome at a desired site. As shown in Figure 4a, we targeted the *c-myc* gene promoter of the chromosome 8 in human genome. The 8-mer G-rich PNA (PNA1 + P) is able to form an invasion complex with the strand in the *c-myc* gene promoter (C42021726A42021752 regions of the chromosome 8 in the human genome). The DNA strands in the complex are susceptible to the catalytic Ce^{IV} /EDTA species and induce a double-strand break in the G-rich sequence (Figure 4a). Two fragments from the scission and *EcoRV* should be observed (2.5 kbp and 7.3 kbp). The whole human genomic DNA as substrate was isolated from cultured HeLa cells. To this human genome, PNA1 + P was added and the mixture was incubated at 37 °C to form an invasion complex. Then, the Ce^{IV} /EDTA complex was added and the scission reaction was carried out at 50 °C for 16 h. The mixture was further digested by *EcoRV*, after being desalted by a spin column. Finally, the products were analyzed by Southern blotting using Probe 1 and 2 (Supporting Information, Table S3) that were synthesized using the PCR DIG probe synthesis kit. As shown in Figure 4a and b, Probe 1 hybridizes to the upstream of the scission site, whereas Probe 2 hybridizes to its downstream. In the Southern blotting using each of these probes, two bands of the expected sizes (2.5 kbp for the upstream fragment and 7.3 kbp for the downstream fragment), which correspond to the dual scission by the present method and *EcoRV*, were explicitly observed. It has been concretely confirmed that the human genomic DNA was specifically cut by the present method.

In summary, this is the first study demonstrating that a single PNA can effectively target the G-rich duplex DNA of the *c-myc* promoter region by formation of a PNA/DNA hybrid G-quadruplex. These results suggest that unmodified PNA has the potential to recognize both strands of G-rich dsDNA in homologous and complementary manners, providing a simple approach to resolve technical difficulties in targeting G-rich sequences by PNA. We have shown that the method was successfully applied to sequence-specific DNA-breaking in the target sequence of human genome. Moreover, it may serve as a new method for the design of more efficient cancer-related G-rich sequence-specific cleaving reagents.

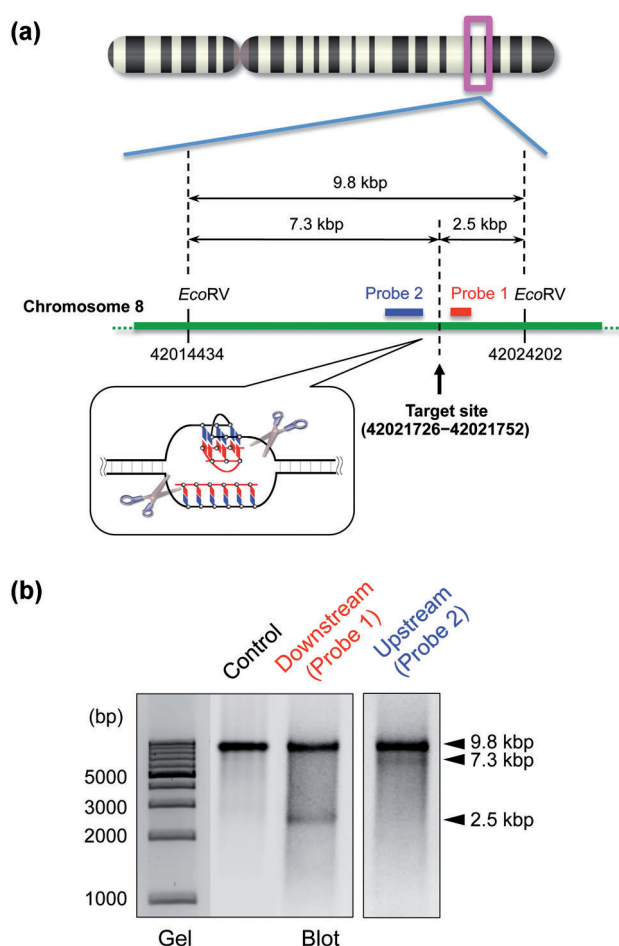


Figure 4. Sequence-specific DNA-breaking of the human genome. a) Sequence-specific scission at the *c-myc* gene promoter of the chromosome 8. After the whole genomic DNA was treated with the PNA1 + P, the product was digested with *EcoRV* and then analyzed by Southern blotting using Probes 1 and 2. Probe 1 and 2 bind to the upstream and the downstream of the *c-myc* gene promoter as a target site, respectively. b) Gel electrophoresis for the G-rich sequence-specific scission of human genomic DNA. The product of G-rich sequence-specific scission was further treated with *EcoRV*, and the whole product was analyzed by Southern blotting. The bands at 2.5 kbp and 7.3 kbp correspond to the DNA fragments formed by the dual scission by the G-rich sequence-specific scission and *EcoRV*, whereas the band at 9.8 kbp comes from the dual scission by *EcoRV* alone (without the scission by the G-rich sequence-specific scission). The markers were stained with GelStar before the specimen was transferred to the blotting membrane. In the control lanes, the human genomic DNA was directly digested by *EcoRV* (without the prior G-rich sequence-specific scission treatment) and subjected to Southern blotting. Reaction conditions: [human genomic DNA] = 40 ng μL^{-1} , [Ce^{IV} /EDTA] = 1 mM, [PNA1 + P] = 500 nM, [HEPES] = 5 mM, and [NaCl] = 10 mM at pH 7.0 and 50 °C for 16 h. To minimize experimental errors, gel loading, incubation time, transfection time of membrane and hybridization time of probe are optimized to assure equality.

Keywords: DNA recognition · DNA scission · G-quadruplex · peptide nucleic acids

Received: February 13, 2012
Revised: April 12, 2012
Published online: June 14, 2012

[1] R. J. Roberts, *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 5905–5908.
[2] A. Pingoud, W. Wende, *ChemBioChem* **2011**, *12*, 1495–1500.

- [3] a) T. Cathomen, J. K. Joung, *Mol. Ther.* **2008**, *16*, 1200–1207; b) F. D. Urnov, E. J. Rebar, M. C. Holmes, H. S. Zhang, P. D. Gregory, *Nat. Rev. Genet.* **2010**, *11*, 636–646.
- [4] a) M. Komiyama, Y. Aiba, Y. Yamamoto, J. Sumaoka, *Nat. Protoc.* **2008**, *3*, 655–662; b) Y. Aiba, J. Sumaoka, M. Komiyama, *Chem. Soc. Rev.* **2011**, *40*, 5657–5668.
- [5] a) J. Lohse, O. Dahl, P. E. Nielsen, *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 11804–11808; b) M. Komiyama, Y. Aiba, T. Ishizuka, J. Sumaoka, *Nat. Protoc.* **2008**, *3*, 646–654; c) T. Ishizuka, J. Yoshida, Y. Yamamoto, J. Sumaoka, T. Tedeschi, R. Corradini, S. Sforza, M. Komiyama, *Nucleic Acids Res.* **2008**, *36*, 1464–1471.
- [6] a) N. Maizels, *Nat. Struct. Mol. Biol.* **2006**, *13*, 1055–1059; b) D. J. Patel, A. T. Phan, V. Kuryavyy, *Nucleic Acids Res.* **2007**, *35*, 7429–7455; c) S. Neidle, *Curr. Opin. Struct. Biol.* **2009**, *19*, 239–250; d) J. Dai, M. Carver, D. Yang, *Biochimie* **2008**, *90*, 1172–1183.
- [7] a) A. K. Todd, M. Johnston, S. Neidle, *Nucleic Acids Res.* **2005**, *33*, 2901–2907; b) J. L. Huppert, S. Balasubramanian, *Nucleic Acids Res.* **2005**, *33*, 2908–2916.
- [8] a) T. Simonsson, P. Pecinka, M. Kubista, *Nucleic Acids Res.* **1998**, *26*, 1167–1172; b) T. A. Brooks, L. H. Hurley, *Nat. Rev. Cancer* **2009**, *9*, 849–861; c) A. T. Phan, Y. S. Modi, D. J. Patel, *J. Am. Chem. Soc.* **2004**, *126*, 8710–8716; d) A. Siddiqui-Jain, C. L. Grand, D. J. Bearss, L. H. Hurley, *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 11593–11598.
- [9] a) S. Balasubramanian, S. Neidle, *Curr. Opin. Chem. Biol.* **2009**, *13*, 345–353; b) T. M. Ou, Y. J. Lu, J. H. Tan, Z. S. Huang, K. Y. Wong, L. Q. Gu, *ChemMedChem* **2008**, *3*, 690–713; c) J. L. Huppert, *Chem. Soc. Rev.* **2008**, *37*, 1375–1384; d) J. L. Mergny, J. F. Riou, P. Mailliet, M. P. Teulade-Fichou, E. Gilson, *Nucleic Acids Res.* **2002**, *30*, 839–865.
- [10] a) J. I. Yeh, B. Shivachev, S. Rapireddy, M. J. Crawford, R. R. Gil, S. Du, M. Madrid, D. H. Ly, *J. Am. Chem. Soc.* **2010**, *132*, 10717–10727; b) M. I. Onyshchenko, T. I. Gaynutdinov, E. A. Englund, D. H. Appella, R. D. Neumann, I. G. Panyutin, *Nucleic Acids Res.* **2009**, *37*, 7570–7580; c) S. Lusvarghi, C. T. Murphy, S. Roy, F. A. Tanious, I. Sacui, W. D. Wilson, D. H. Ly, B. A. Armitage, *J. Am. Chem. Soc.* **2009**, *131*, 18415–18424; d) G. F. He, S. Rapireddy, R. Bahal, B. Sahu, D. H. Ly, *J. Am. Chem. Soc.* **2009**, *131*, 12088–12090; e) A. Paul, P. Sengupta, Y. Krishnan, S. Ladame, *Chem. Eur. J.* **2008**, *14*, 8682–8689; f) S. Roy, F. A. Tanious, W. D. Wilson, D. H. Ly, B. A. Armitage, *Biochemistry* **2007**, *46*, 10433–10443; g) V. L. Marin, B. A. Armitage, *Biochemistry* **2006**, *45*, 1745–1754; h) E. A. Englund, Q. Xu, M. A. Witschi, D. H. Appella, *J. Am. Chem. Soc.* **2006**, *128*, 16456–16457; i) J. Seenisamy, S. Bashyam, V. Gokhale, H. Vankayalapati, D. Sun, A. Siddiqui-Jain, N. Streiner, K. Shin-ya, E. White, W. D. Wilson, L. H. Hurley, *J. Am. Chem. Soc.* **2005**, *127*, 2944–2959; j) A. T. Phan, V. Kuryavyy, H. Y. Gaw, D. J. Patel, *Nat. Chem. Biol.* **2005**, *1*, 167–173; k) B. Datta, B. A. Armitage, *J. Am. Chem. Soc.* **2001**, *123*, 9612–9619.
- [11] T. Ishizuka, K. Otani, J. Sumaoka, M. Komiyama, *Chem. Commun.* **2009**, 1225–1227.
- [12] a) L. A. Cahoon, H. S. Seifert, *Science* **2009**, *325*, 764–767; b) J. Tang, Z. Y. Kan, Y. Yao, Q. Wang, Y. H. Hao, Z. Tan, *Nucleic Acids Res.* **2008**, *36*, 1200–1208; c) K. Guo, A. Pourpak, K. Beetz-Rogers, V. Gokhale, D. Sun, L. H. Hurley, *J. Am. Chem. Soc.* **2007**, *129*, 10220–10228.
- [13] a) Y. Xue, J. Q. Liu, K. W. Zheng, Z. Y. Kan, Y. H. Hao, Z. Tan, *Angew. Chem.* **2011**, *123*, 8196–8200; *Angew. Chem. Int. Ed.* **2011**, *50*, 8046–8050; b) Y. Xu, *Chem. Soc. Rev.* **2011**, *40*, 2719–2740; c) Y. Xu, Y. Noguchi, H. Sugiyama, *Bioorg. Med. Chem.* **2006**, *14*, 5584–5591; d) B. Datta, C. Schmitt, B. A. Armitage, *J. Am. Chem. Soc.* **2003**, *125*, 4111–4118.
- [14] Y. Xu, Y. Suzuki, T. Lönnberg, M. Komiyama, *J. Am. Chem. Soc.* **2009**, *131*, 2871–2874.