Sequence-Specific Recognition of Double-Stranded DNA by Using Only PNAs in Parallel with Natural Nucleobases

Masanari Shibata, Yuichiro Aiba*, Masaki Hibino, and Osami Shoji*

Department of Chemistry, Graduate School of Science, Nagoya University, Furo-cho, Chikusa-ku, Nagoya, 464-8602 (Japan)

*To whom correspondence should be addressed.

E-mail: aiba.yuichiro.f4@f.mail.nagoya-u.ac.jp, shoji.osami.w3@f.mail.nagoya-u.ac.jp

Abstract

The sequence-specific recognition of double-stranded DNA (dsDNA) is a key property for the control of DNA function. Peptide nucleic acid (PNA) can be utilised for the direct recognition of dsDNA via the formation of a unique invasion complex. Strand invasion by PNA induces local changes in the structure of dsDNA and is useful for the regulation of gene expression and genome editing. However, the fact that nucleobases modification is required for efficient invasion, has stymied the wide-spread application of PNA. Herein, we succeeded in the efficient recognition of target dsDNA sequences via formation of invasion complex by utilising only parallel-stranded and unmodified PNAs. This approach also streamlines synthesis by permitting the use of a peptide synthesiser rather than the manual synthesis we had been dependent upon for nucleobase-modified PNAs. Our new method also exhibited high sequence specificity and flexibility for target dsDNA sequences.

Introduction

The sequence-specific recognition of DNA has been extensively studied owing to its wide range of potential applications. Nucleic acids themselves have been good candidates for the recognition and control of target DNA by nature of their complementarity, enabling, for example, RNA/DNA duplexes to stably form A-/B-type double helices based on standard Watson-Crick base-pairing rules.[1] In addition to the natural nucleic acids DNA and RNA, a variety of artificial nucleic acids (XNAs from xeno nucleic acids) with modified nucleobases, backbones, or a combination of both have been developed for the purpose of bestowing nucleic acids with novel properties and functions.[2-15] Such XNAs are also powerful tools for the recognition of DNA/RNA sequences and the formation of unique structures. [14-15]

Peptide nucleic acid (PNA) is a type of synthetic DNA analogue, wherein the negatively charged sugar-phosphate backbone has been replaced by an electrostatically neutral pseudopeptide backbone (Figure 1a). [16] This has to consequence that there is no electrostatic repulsion between PNA and the phosphate backbone of DNA, meaning PNA can form a more stable duplex with DNA than the equivalent DNA/DNA duplex. [17] Moreover, PNA can form an atypical and unique invasion complex by penetrating into double-stranded DNA (dsDNA), a phenomenon that has not been reproduced in the same way with other artificial nucleic acids. [18-21] As depicted in Figure 1b, two PNA strands, complementary to each strand of the target DNA duplex, can invade into the DNA double helix to two separate anti-parallel PNA/DNA duplexes. This unique structure is called a double-duplex invasion complex or, more simply, an invasion complex. As PNA forms invasion complexes sequence-selectively, inducing significant changes in the local structure of target dsDNA without requiring pre-denaturation, PNA is anticipated to be a powerful tool for the control of DNA function *in cellula* and *in vivo*. [22]

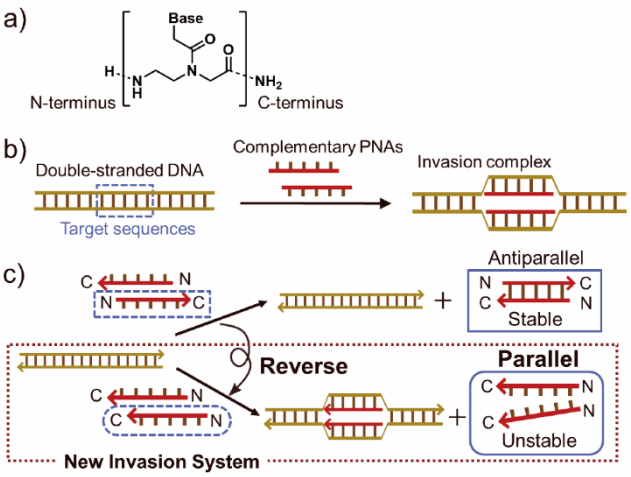

Figure 1. (a) Chemical structure of peptide nucleic acid (PNA). (b) Unique DNA recognition of PNA

via invasion complex formation. (c) Our new parallel-stranded PNA invasion system: The use of parallel-stranded PNAs inhibits undesirable PNA duplex formation, and enables the formation of invasion complex without the need for chemically modified PNAs. N and C stand for the N- and C-termini of PNA, respectively.

To be precise, however, a trick is required for double-duplex invasion to even take place. This is due to the two PNA strands used for invasion being inevitably complementary, meaning that an anti-parallel PNA/PNA duplex, which is more stable than the corresponding PNA/DNA duplex, is preferentially formed. To circumvent this problem, pseudo-complementary PNA (pcPNA) has been developed, where the adenine (A) and thymine (T) nucleobases have been replaced by 2,6-diaminopurine (D) and 2-thiouracil (Us), respectively.[18, 23-25] This introduces steric repulsion between the amino group of D and the thione group of Us, which significantly destabilises the formation of duplexes between complementary strands of pcPNA (Figure S1). However, there is no steric repulsion between pcPNAs and natural nucleobases, so that a pair of stable pcPNA/DNA duplexes can form via the invasion of pcPNAs into dsDNA. Although effective invasion with pcPNA has been reported, the synthesis of pcPNA monomers is costly and time-consuming. Moreover, pcPNA monomers are commercially available exclusively as Boc-protected monomers, which renders automated synthesis challenging, consequently hampering the large-scale application of pcPNA oligomers. However, to realise a diverse range of applications, it is imperative to overcome this limitation, that is, the requirement for modified PNA nucleobases to achieve efficient invasion.

Herein, we propose a simple yet elegant solution to this predicament, achieving efficient invasion using only basic unmodified PNAs, and evaluated their applicability for the sequence-selective recognition of dsDNA. This “novel” approach to suppress the formation of PNA duplexes focused on a phenomenon that has largely been overlooked, namely the differences in duplex stability depending upon the relative strand orientation. Owing to the asymmetric and oriented backbone structure of DNA, RNA, and XNAs including PNA, two distinct kinds of binding orientation, namely parallel and antiparallel, can be formed, even if the sequences are identical. In nature, dsDNA generally exists in its antiparallel form, wherein two DNA strands are aligned in opposite directions with the 5'-end of one stand facing the 3'-end of the other strand. Analogously, a strand of PNA possesses an N- and C-terminus, which correspond to the 5'- and 3'-end of DNA, respectively, and can form both stable antiparallel and parallel-stranded duplexes with complementary DNA.[17] However, a significant difference in thermal stability exists between antiparallel- and parallel-stranded PNA duplexes, with the parallel-stranded PNA duplex being much less stable than the corresponding antiparallel-stranded PNA duplex.[26] Based upon these reported observations, we concluded that by designing two parallel-strands of PNA, as shown in Figure 1, the formation of an undesirable PNA duplex should be significantly suppressed, greatly encouraging invasion into dsDNA. If this rationale were shown to hold true, invasion into dsDNA should now be possible without the need for modification of PNA

monomers.

Results and Discussion

PNA (not pcPNA) oligomers were synthesised on an automated solid-phase peptide synthesiser (Syro I, Biotage) using Fmoc-PNA monomers. All PNA oligomers possessed a free N-terminal amino group and a C-terminal carboxylic acid that was converted to an amide. Twelve PNA oligomers were designed, with their sequences shown in Table 1 and Table S1. The formation of invasion complex was initiated by mixing a pair of PNAs with target 119-bp dsDNA, and the efficiency of invasion complex formation was evaluated by electrophoretic mobility shift assay (EMSA) using a microtip electrophoresis system (MCE-202 MultiNA, Shimadzu). The invasion complex can be observed as a band with a lower electrophoretic mobility, arising from changes in the local structure of dsDNA at the invasion site.[27] Henceforth, PNAs will be referred to as either antiparallel (apsPNA) or parallel stranded (psPNA) with respect to the strand orientation of the resulting PNA duplex formed with PNA-Fw. Consequently, PNA/DNA duplexes formed between DNA and a strand of either PNA-Fw or apsPNAs form antiparallel PNA/DNA duplexes, whereas duplexes formed between DNA and a strand of psPNAs form parallel-stranded PNA/DNA duplexes (Figure 2a).

Table 1. Sequences of PNAs and DNAs.

Name†	Sequences of PNAs (N to C) and DNAs (5' to 3')†
PNA-Fw	*K*ATGACTAAGAGTAGT*KK*
apsPNA-Rev	*K*ACTACTCTTAGTCAT*KK*
psPNA-Rev	*K*TACTGATTCTCATCA*KK*
PNA-Fw-mis	*K*ATGACTATGAGTAGT*KK*
psPNA-Rev-mis	*K*TACTGATACTCATCA*KK*
PNA-Fw-GC	*K*GGTAGCGGCTGAAGC*KK*
apsPNA-Rev-GC	*K*GCTTCAGCCGCTACC*KK*
psPNA-Rev-GC	*K*CCATCGCCGACTTCG*KK*
DNA-Fw	ATGACTAAGAGTAGT
DNA-Rev	ACTACTCTTAGTCAT

† K = lysine, aps = antiparallel stranded, ps = parallel stranded; aps and ps are given with respect to the strand orientation of the duplex formed with the complementary PNA-Fw strand.

Adding of PNA-Fw and apsPNA-Rev to a solution containing target 119-bp dsDNA only yielded a single band corresponding to the target dsDNA. Additionally, no formation of invasion complex could be observed, even when the equivalents of PNA to the target dsDNA were increased (Figure 2b; lanes 3 – 5). This clearly shows that, in the absence of modified nucleobases, i.e. pcPNA, a pair of antiparallel-stranded PNAs prefer the formation of the more stable antiparallel-stranded PNA duplex to the corresponding PNA/DNA duplexes. Consequently, the effective concentration of PNA available

for invasion is decreased and no invasion complex is formed. On the other hand, when PNA-Fw and psPNA-Rev, forming a less stable parallel-stranded PNA duplex, were employed, a new lower-mobility band was observed. Moreover, the band became more intense with increasing equivalents of PNA (Figure 2b; lanes 6 – 8). Interestingly, this new band was not observed when only one of the two PNA strands was mixed with target dsDNA, meaning that this recognition system requires the cooperative action of both PNA-Fw and psPNA-Rev with dsDNA for invasion to proceed. These results allow us to conclude that our newly developed parallel-stranded PNAs interact with dsDNA as intended, and that the new band is indeed the desired invasion complex.

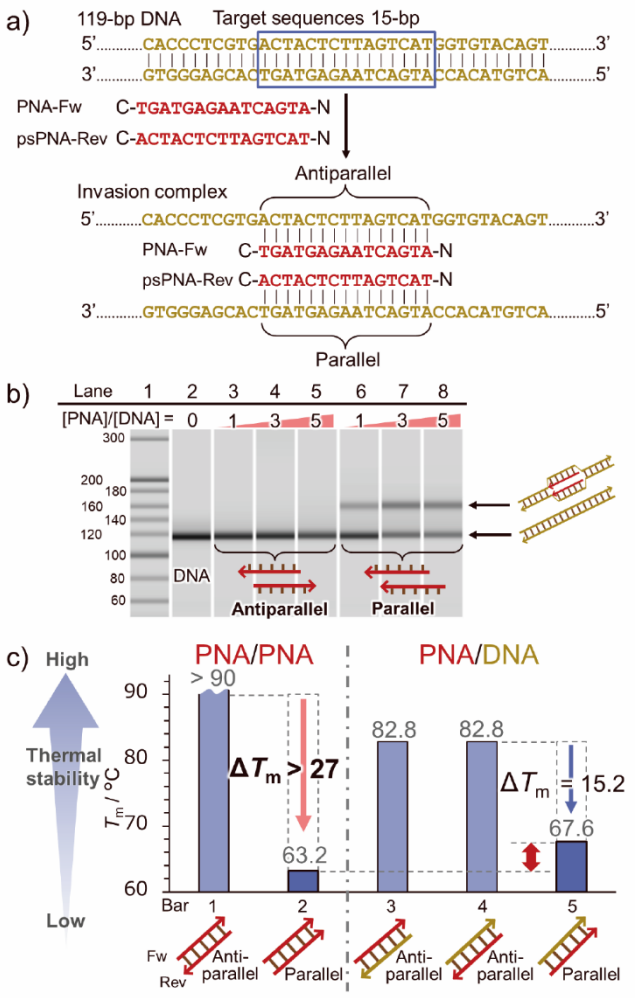

Figure 2. (a) Overview of invasion experiment with parallel-stranded PNAs. (b) Electrophoretic mobility shift assay (EMSA) showing that parallel-stranded PNA pairs form an invasion complex. Lane1: 20-bp DNA Ladder; lane 2: 119-bp target DNA only; lanes 3 – 5: PNA-Fw and apsPNA-Rev; lanes 6 – 8: PNA-Fw and psPNA-Rev. Invasion conditions: [DNA] = 100 nM, [each PNA] = 100 – 500 nM (1 – 5 eq. for DNA), and [HEPES (pH 7.0)] = 5 mM at 50 °C for 1 h. (c) Melting temperature (Tm) of PNA/PNA (bars 1, 2) and PNA/DNA (bars 3 – 5) duplexes for evaluation of the thermal

stability of PNA and DNA duplexes. Tm values are shown above the corresponding bars. Bar 1, PNA-Fw/apsPNA-Rev; bar 2, PNA-Fw/psPNA-Rev; bar 3, PNA-Fw/DNA-Rev; bar 4, DNA-Fw/apsPNA-Rev; bar 5, DNA-Fw/psPNA-Rev, Conditions: [each strand of DNA or PNA] = 1 μM, and [HEPES (pH 7.0)] = 5 mM.

To better understand the mechanism governing our new parallel-stranded PNA invasion system, we evaluated the thermal stability of antiparallel- and parallel-stranded PNA/PNA and PNA/DNA duplexes by measuring their melting temperature (Tm) values. The results are summarised in Figure 2c. As expected, a very low Tm value was obtained for the parallel-stranded PNA duplex (PNA-Fw/psPNA-Rev; 63 °C) as compared with the antiparallel-stranded PNA duplex (PNA-Fw/apsPNA-Rev; over 90 °C). Regarding PNA/DNA duplexes, all antiparallel- (DNA-Fw/apsPNA-Rev; 83 °C, PNA-Fw/DNA-Rev; 83 °C) and parallel-stranded (psPNA-Fw/DNA-Rev; 68 °C) duplexes exhibit much higher Tm values than that of the corresponding DNA/DNA duplex (DNA-Fw/DNA-Rev; 31 °C). Moreover, both PNA/DNA duplexes related to invasion complex formation (83 & 68 °C) are more stable than the corresponding parallel-stranded PNA duplex (PNA-Fw/psPNA-Rev; 63 °C). Accordingly, we concluded that it is the difference in thermal stability between PNA/PNA and PNA/DNA duplexes, especially the increased stability of parallel-stranded PNA/DNA duplexes over parallel-stranded PNA duplexes, that drives the favourable formation of the target invasion complex.

To test the limits of this new invasion system, we investigated I) shorter PNAs, II) sequence specificity, and III) sequence flexibility. As for the relationship between invasion efficiency and the length of the PNA oligomer, shorter 11-mer and 12-mer parallel-stranded PNAs were synthesised on a peptide synthesiser according to the same method described for 15-mer PNAs and evaluated by EMSA (Figure S2, Table S1). Whilst efficiency was significantly lower than for 15-mer PNAs (lane 3 – 5), formation of invasion complex was possible even with the shorter parallel-stranded PNAs, which is promising for the targeting of shorter sequences. Subsequently, sequence selectivity of parallel-stranded PNAs was investigated using 15-mer PNAs. Sequence-specific recognition of dsDNA is a crucial point for the practical application of our method. A point mutation was introduced at the centre of the parallel-stranded PNA pair to yield a single base-pair mismatch at the position highlighted in Figure 3a. Parallel-stranded PNAs exhibit high invasion efficiency for fully matching target dsDNA (lanes 3 – 4), however, following introduction of a single base-pair mismatch into the PNA, no band corresponding to invasion complex could be detected, even after the equivalents of PNA were increased (lanes 5 – 6). These results exemplified that parallel-stranded PNAs recognise dsDNA with high sequence selectivity.

Finally, we examined whether our parallel-stranded PNA system is valid for different target sequences. From a practical viewpoint, it is crucial that parallel-stranded PNAs can be applied to a wide variety

of target sequences of dsDNA. Thus, we designed parallel-stranded PNAs (PNA-Fw-GC & PNA-Rev-GC) targeting a G-C rich sequence with a completely different nucleotide composition from that of the A-T rich one (Figure 2). As presented in Figure 3b, our newly synthesised G-C rich parallel-stranded PNAs could also accurately recognise the target dsDNA and form the respective invasion complex, demonstrating the ability of the parallel-stranded PNA system to target both A-T and G-C rich sequences. Whilst, theoretically, the proportion of modified nucleobases is an important factor dictating the efficiency of invasion for pcPNAs, this is not the case for parallel-stranded PNAs. Although the invasion to GC-rich sequence was slightly less efficient than that to AT-rich sequence, this result greatly expands the range of double-duplex invasion applications in terms of the ability to appropriately recognise sequences with widely different nucleobase compositions.

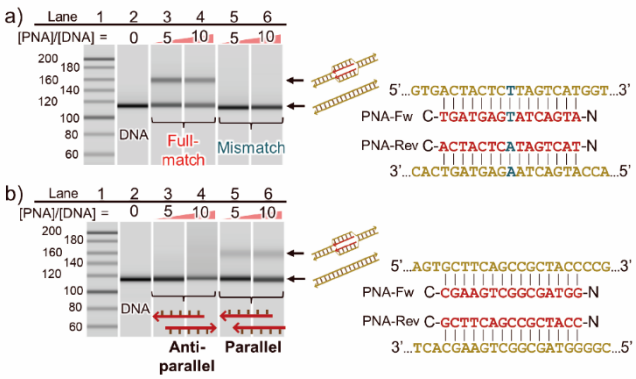

Figure 3. (a) EMSA for evaluation of the sequence selectivity of the parallel invasion system. Lane 1: 20-bp DNA Ladder; lane 2: 119-bp DNA only; lanes 3 and 4: PNA-Fw and psPNA-Rev; lanes 5 and 6: PNA-Fw-mis and psPNA-Rev-mis. (b) EMSA demonstrating that the parallel invasion system is adaptable to other DNA sequences. Lane 1: 20-bp DNA Ladder; lane 2: 119-bp DNA only; lanes 3 and 4: PNA-Fw-GC and apsPNA-Rev-GC; lanes 5 and 6: PNA-Fw-GC and psPNA-Rev-GC. Invasion conditions: [DNA] = 100 nM, [each PNA] = 500 - 1000 nM (5 - 10 eq. for DNA), and [HEPES (pH 7.0)] = 5 mM at 50 °C for 1 h.

Conclusion

Via a simple and elegant approach employing plain, unmodified PNAs in parallel, we have successfully constructed an extremely simple dsDNA recognition system based upon invasion by PNA, removing the requirement for the laborious and costly modification of nucleobases. This brings us a significant step closer to an "ideal" invasion system, which has been considered impossible without the use of synthetically complex PNA derivatives. Through the use of parallel-stranded PNA, this belief could be upended, and the site-selective invasion of simple unmodified PNA into dsDNA was achieved. PNA oligomers with natural nucleobases are easily synthesised by combining commercially available Fmoc PNA monomers and automated peptide synthesisers, unlike manually synthesised

pcPNAs containing modified nucleobases.[18, 23-25] In fact, all PNAs used in this study were prepared by automated synthesis with a peptide synthesiser. PNAs without any nucleobase modification for A/T are more preferable in terms of cost as well as the range of targetable sequences (applicability to GC-rich sequences). We believe that elimination of our dependence on pcPNA will undoubtedly lead to a great leap forward in the research of PNA invasion.

Furthermore, in our system, it is easy to functionalise PNAs by employing bioconjugation techniques for peptides[28-29] to introduce various functional molecules such as cell-penetrating peptides,[28, 30] nuclear localisation signal peptides,[31-32] peptide tags, dyes,[33-34] and metallocomplexes[35]. Although there is still room for improvement regarding invasion efficiency, it is expected that the parallel-stranded PNA system can be applied in various ways including in gene regulation as well as genome editing by applying the accumulated knowledge of PNA chemistry.

References

[1] R. Dickerson, H. Drew, B. Conner, R. Wing, A. Fratini, M. Kopka, *Science* **1982**, *216*, 475-485.

[2] S. Obika, D. Nanbu, Y. Hari, K.-I. Morio, Y. In, T. Ishida, T. Imanishi, *Tetrahedron Lett.* **1997**, *38*, 8735-8738.

[3] M. D'Costa, V. A. Kumar, K. N. Ganesh, *Org. Lett.* **1999**, *1*, 1513-1516.

[4] T. Wada, N. Minamimoto, Y. Inaki, Y. Inoue, *J. Am. Chem. Soc.* **2000**, *122*, 6900-6910.

[5] N. Oka, T. Wada, K. Saigo, *J. Am. Chem. Soc.* **2003**, *125*, 8307-8317.

[6] L. Zhang, A. Peritz, E. Meggers, *J. Am. Chem. Soc.* **2005**, *127*, 4174-4175.

[7] A. Dragulescu-Andrasi, S. Rapireddy, B. M. Frezza, C. Gayathri, R. R. Gil, D. H. Ly, *J. Am. Chem. Soc.* **2006**, *128*, 10258-10267.

[8] T. S. Kumar, A. S. Madsen, J. Wengel, P. J. Hrdlicka, *J. Org. Chem.* **2006**, *71*, 4188-4201.

[9] H. Kashida, K. Murayama, T. Toda, H. Asanuma, *Angew. Chem. Int. Ed.* **2011**, *50*, 1285-1288.

[10] P. Sriwarom, P. Padungros, T. Vilaivan, *J. Org. Chem.* **2015**, *80*, 7058-7065.

[11] D. U. Ukale, T. Lönnberg, *Angew. Chem. Int. Ed.* **2018**, *57*, 16171-16175.

[12] M. Hibino, Y. Aiba, O. Shoji, *Chem. Commun.* **2020**, *56*, 2546-2549.

[13] Y. Takezawa, A. Suzuki, M. Nakaya, K. Nishiyama, M. Shionoya, *J. Am. Chem. Soc.* **2020**, *142*, 21640-21644.

[14] D. Praseuth, A. L. Guieysse, C. Hélène, *Biochim. Biophys. Acta, Gene Struct. Expression* **1999**, *1489*, 181-206.

[15] H. Kashida, Y. Hattori, K. Tazoe, T. Inoue, K. Nishikawa, K. Ishii, S. Uchiyama, H. Yamashita, M. Abe, Y. Kamiya, H. Asanuma, *J. Am. Chem. Soc.* **2018**, *140*, 8456-8462.

[16] P. E. Nielsen, M. Egholm, R. H. Berg, O. Buchardt, *Science* **1991**, *254*, 1497-1500.

[17] M. Egholm, O. Buchardt, L. Christensen, C. Behrens, S. M. Freier, D. A. Driver, R. H. Berg, S. K. Kim, B. Norden, P. E. Nielsen, *Nature* **1993**, *365*, 566-568.

[18] J. Lohse, O. Dahl, P. E. Nielsen, *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 11804-11808.

[19] Recent studies have reported the formation of non-PNA invasion complexes using artificial nucleic acids with well-designed special functions (see Ref. 20 and 21).

[20] S. Nakamura, H. Kawabata, K. Fujimoto, *Chem. Commun.* **2017**, *53*, 7616-7619.

[21] R. G. Emehiser, P. J. Hrdlicka, *Org. Biomol. Chem.* **2020**, *18*, 1359-1368.

[22] A. S. Ricciardi, E. Quijano, R. Putman, W. M. Saltzman, P. M. Glazer, *Molecules* **2018**, *23*.

[23] L. Christensen, R. Fitzpatrick, B. Gildea, K. H. Petersen, H. F. Hansen, T. Koch, M. Egholm, O. Buchardt, P. E. Nielsen, J. Coull, R. H. Berg, *J. Pept. Sci.* **1995**, *1*, 175-183.

[24] G. Haaima, H. F. Hansen, L. Christensen, O. Dahl, P. E. Nielsen, *Nucleic Acids Res.* **1997**, *25*, 4639-4643.

[25] M. Komiyama, Y. Aiba, T. Ishizuka, J. Sumaoka, *Nat. Protoc.* **2008**, *3*, 646-654.

[26] P. Wittung, P. E. Nielsen, O. Buchardt, M. Egholm, B. Norden, *Nature* **1994**, *368*, 561-563.

[27] H. Kuhn, D. I. Cherny, V. V. Demidov, M. D. Frank-Kamenetskii, *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 7548-7553.

[28] M. Pooga, U. Soomets, M. Hällbrink, A. Valkna, K. Saar, K. Rezaei, U. Kahl, J.-X. Hao, X.-J. Xu, Z. Wiesenfeld-Hallin, T. Hökfelt, T. Bartfai, Ü. Langel, *Nat. Biotechnol.* **1998**, *16*, 857-861.

[29] L. Good, S. K. Awasthi, R. Dryselius, O. Larsson, P. E. Nielsen, *Nat. Biotechnol.* **2001**, *19*, 360-364.

[30] J. J. Turner, G. D. Ivanova, B. Verbeure, D. Williams, A. A. Arzumanov, S. Abes, B. Lebleu, M. J. Gait, *Nucleic Acids Res.* **2005**, *33*, 6837-6849.

[31] Y. Aiba, Y. Honda, M. Komiyama, *Chem. Eur. J.* **2015**, *21*, 4021-4026.

[32] Y. Aiba, G. Urbina, M. Shibata, O. Shoji, *Appl. Sci.* **2020**, *10*, 8663.

[33] N. Svanvik, J. Nygren, G. Westman, M. Kubista, *J. Am. Chem. Soc.* **2001**, *123*, 803-809.

[34] O. Seitz, *Angew. Chem. Int. Ed.* **2000**, *39*, 3249-3252.

[35] M. Hibino, Y. Aiba, Y. Watanabe, O. Shoji, *ChemBioChem* **2018**, *19*, 1601-1604.

Acknowledgement

This research was partially supported by JSPS KAKENHI (grant number 19K05730 and 22K05350 to Y.A., and 22J14529 to M.S.) from the Ministry of Education, Culture, Sports, Science, and Technology (Japan), Foundation of Public Interest of Tatematsu (Y.A.), and the Noguchi Institute (Y.A.). We thank Dr Joshua Kyle Stanfield for checking the English of this manuscript.

Author contributions

M.S. and Y.A. conceived the project. M.S. performed purification of PNA oligomers, electrophoretic mobility shift assay and melting temperature analysis with the help of M.H. M.S. wrote the first draft of the manuscript. M.S., Y.A., and O.S. prepared the manuscript. All authors discussed the manuscript and approved the final version.