DNA, PNA, and Their Derivatives for Precise Genotyping of SNPs

A. Kuzuya, J.-M. Zhou and M. Komiyama^{*}

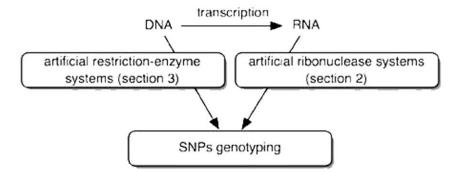
Research Center for Advanced Science and Technology, The University of Tokyo, 4-6-1 Komaba, Meguro-ku, Tokyo, 153-8904 Japan.

Abstract: Two new approaches for SNP genotyping are described. The one is based on tandem site-selective RNA scission and the genotype is determined by MALDI-TOF/MS analyses of clipped short RNA. The other is visual SNP genotyping with combinations of peptide nucleic acid (PNA) and single-stranded DNA specific nucleases.

Keywords: SNP genotyping, RNA, PNA, selective scission, acridine, nuclease.

1. INTRODUCTION

High-throughput, accurate, and low-cost methods for genotyping of single nucleotide polymorphisms (SNPs) [1] are crucially important for design of tailor-made medicines, prediction of hereditary diseases, and many other biomedical applications. Although a number of methods were hitherto proposed, few of them are satisfactory. We recently developed two methods for this purpose, which will be The second method is to combine peptide nucleic acid (PNA) with single-stranded DNA specific nucleases (i.e. nuclease S1, nuclease P1 and Mung Bean nuclease) [4]. We found that only the DNA in DNA/PNA duplexes involving a mismatch are efficiently hydrolyzed by these nucleases, whereas fully-matching sequences are kept intact (artificial restriction-enzyme systems). This difference is visually scored by adding 3,3'-diethylthiadicarbocyanine, which



Scheme 1. The two strategies for precise SNP genotyping.

described in this review (see Scheme (1)). The first method is to use modified oligonucleotides for selective RNA activation at two sites and clip appropriate RNA fragments from human gene [2]. It is based on our recent finding of site-selective RNA scission that involves no covalent fixation of catalysts to sequence-recognizing moiety (artificial ribonuclease systems) [3]. By non-covalent interactions of the RNA with oligonucleotide bearing two acridine groups, the phosphodiester linkages in front of these acridine groups are selectively activated and hydrolyzed by various catalysts (e.g. lanthanide(III) ions). Advantageously, it requires no specific base-sequence at the scission site. By analyzing the RNA fragments by MALDI-TOF MS, any alternation at the target nucleotide is definitely pinned down. changes its color from blue to purple upon binding to DNA/PNA duplexes. These findings are applied to convenient and straightforward detection of SNPs. When the target site in the sample DNA is completely complementary with the PNA, notable amount of DNA/PNA duplex remains, and thus the solution exhibits purple color. In the presence of even one mismatch between PNA and DNA, however, the DNA is completely digested by the enzyme and therefore, the dye shows its intrinsic blue color. The SNPs in the apolipoprotein E gene of human DNA have been successfully genotyped by this method.

2. SNP GENOTYPING WITH THE USE OF ARTIFICIAL RIBONUCLEASE SYSTEMS

2.1 Strategy for Site-Selective RNA Scission

Sequence-selective scission of RNA has been attracting interests of chemists and biochemists, since it is essential for molecular biology and therapy in the future (e.g. regulation of specific gene expression, advanced therapy, and

^{*}Address correspondence to this author at the Research Center for Advanced Science and Technology, The University of Tokyo, 4-6-1 Komaba, Meguro-ku, Tokyo, 153-8904 Japan; E-mail: komiyama@mkomi.rcast.u-tokyo.ac.jp

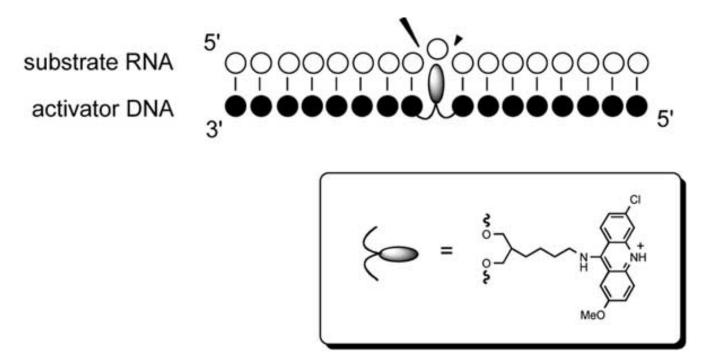


Fig. (1). Schematic representation of the artificial ribonuclease system. Selective scission sites are indicated by wedge-shaped arrows.

RNA manipulation) [5]. Many artificial enzymes were prepared by attaching molecular scissors to sequencerecognizing oligonucleotide. Recently, however, we have developed entirely new artificial ribonuclease systems. The first SNP genotyping method presented in this review is based on these systems, and thus their fundamentals will be first described rather in detail.

These new artificial ribonuclease systems take advantage of selective activation of the target phosphodiester linkage in the RNA by oligonucleotides bearing an acridine as the activator (Fig. 1). When the substrate RNA forms a heteroduplex with this activator, most of the ribonucleotides, except for the one in front of the acridine, form Watson-Crick base pairs with the counterpart nucleotides. As a result, the phosphodiester linkages adjacent to the unpaired nucleotide (indicated by the wedgeshaped arrows) are selectively activated. In contrast, the other phosphodiester linkages in the substrate RNA are deactivated on the formation of these heteroduplexes. Accordingly, these linkages are preferentially hydrolyzed by LuCl3 (molecular scissors) over the other linkages. Thus, the selective scission occurs at the 5'-side of the target nucleotide (major scissionsite) and its 3'-side (minor scission-site). In contrast with previous artificial ribonucleases, covalent attachment of the molecular scissors to any sequence-recognizing moiety is not necessary. The sequence-selectivity for these non-covalent systems is satisfactorily high throughout the whole reaction process. This sequence-selective activation is so efficient that the scission is prompt even with intrinsically poor catalysts, such as Zn(II) or Mn(II) ions. The scission is marginal when a trimethylene spacer without an acridine is inserted to DNA. Apparently, the acridine residue is essential for the present site-selective scission. What is crucial is not the presence of unpaired nucleotide in RNA substrate, but the interactions between the acridine and the RNA leading to the RNA activation.

In all the RNA sequences investigated, the phosphodiester linkage opposite the acridine is activated and selectively hydrolyzed. Thus, the nucleotide at the target site in RNA must remain unpaired for the site-selective scission. When all the nucleotides in RNA are forming Watson-Crick base pairs with the "activators", the selective scission completely disappears. Both the chloro- and methoxy-groups on the acridine are also important (although not essential). In so far as these requirements are met, the scission efficiencies are not much dependent on the target sequence. This is one of the most important advantages of the present artificial ribonuclease systems, since conventional ribozymes usually require specific sequences at the target site. Because of this advantage, SNP genotyping is possible for all the positions of the substrate RNA (*vide infra*).

Spectroscopic analysis (fluorescence, UV, and CD) has shown that, in the DNA/RNA heteroduplexes, the acridine group is sandwiched between the two neighboring base pairs. The acridine is so large in size that it cannot be simultaneously accommodated together with the targetunpaired nucleobase. As a result, the unpaired nucleobase is supposed to be somewhat pushed out, perturbing the conformation of the target site in the RNA. The 2'O atom at the scissile linkage is located near the corresponding phosphorus atom, and its intramolecular nucleophilic attack is promoted. Moreover, general-acid catalysis by the ringnitrogen of the acridine facilitates the release of the leaving 5'O from the pentacoordinated intermediate. [6]

2.2 Application to SNP Genotyping

These artificial ribonuclease systems are simple and active enough to be successfully applied to SNP genotyping. A schematic representation of the analytical procedures is shown in Fig. (2). In order to activate the substrate RNA at two designated sites, two acridines are bound to

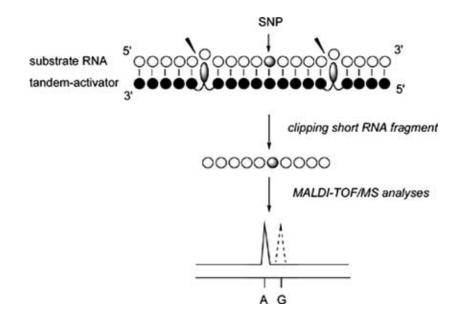


Fig. (2). Flowchart of SNP genotyping by selective RNA clipping. The nucleotide in the SNP site is strictly determined by mass difference of the fragment.

oligonucleotide and this modified DNA is used as the tandem-activator. The two target sites are selected so that the SNP site is placed between them. Upon treating this system with free lanthanide(III) ion, short RNA fragment containing SNP site is selectively clipped from the substrate RNA. By analyzing this fragment with MALDI/TOF-MS, the nucleotide in the SNP site can be precisely determined.

In the following examples, the sample RNAs are the ones coding the exon of human apolipoprotein E gene (apoE) [7]. There are two SNP sites in this gene. The A \rightarrow G alternation at the nucleotide position 152 in its antisense strand (this isoform is apoE4) is responsible for Alzheimer's disease, whereas the $G \rightarrow A$ alternation at the nucleotide position 290 (apoE2) is associated with hyperlipemia. The RNA coding a portion from A170 to C131 in A-allele of apoE is treated with LuCl₃ in the presence of the tandemactivator, which has two acridine groups in front of G158 and U144 (see Fig. 3). The potential SNP site A152 (in bold) is in this sequence. Note that the tandem-activator used here involves non-natural nucleotide dP in front of the SNP site. This nucleotide analog forms stable base pairs with both G and A through its tautomerism. Consequently, this tandem-activator can strongly bind the substrate (and the fragment formed by the two-site scission), irrespective of the genotype of the substrate. As shown in Fig. (3b), MALDI-TOF mass spectrum of the products clearly exhibits a peak at m/z = 4496.8 (in bold), which is fairly in accord with the theoretical value (4496.7) for the 14-mer fragment from G158 to C145. Both of the target sites are efficiently activated by the acridines and hydrolyzed by Lu(III). These facts have been further substantiated by gel electrophoresis, in which the substrate RNA is labeled at both ends by two different dyes. The scission efficiency at each of the designated sites is similar to the value accomplished by the oligonucleotide bearing only one acridine. Since the present artificial ribonuclease systems can cut RNA at any target sequence, as described in section 2.1, the length of the clipped fragment can be freely chosen by using appropriate tandem-activator, in which two acridines are placed at the corresponding distance. Other lanthanide(III) ions (e.g. La(III) and Eu(III)) can be also used as molecular scissors if necessary.

The present genotyping method is also applicable to heterozygous samples. In such samples, the RNA from the subject involves two components that differ from each other in only one nucleobase, corresponding to the two alleles on homologous chromosomes. In Fig. (3c), 1:1 mixture of two substrates corresponding to each alleles is treated with the combination of Lu(III) and the tandem-activator (the DNA bearing two acridine). Non-natural nucleotide dP is used in front of the SNP site. Exactly as designed, two signals of similar intensity are detected at m/z 4510.0 and 4494.1, which correspond to the 14-mer fragment from G-allele (theoretical m/z = 4512.6) and the one from A-allele, respectively. By using appropriate tandem-activator, 1:1 mixtures of two alleles for the other SNP site are also analyzable. The degenerate recognition of both A and G by dP is essential for precise genotyping of heterozygous samples. Thus, no signal for A-allele is detectable when conventional nucleotide C is used in front of the SNP site. Here, the RNA fragment from this allele has a mismatch to the activator and is bound only weakly, so that it is promptly digested by the Lu(III) in the solution.

Multiplex analyses are also possible in the present method. For example, by using two tandem-activators simultaneously, two RNA fragments involving two SNP sites (position 152 and 290) can be successfully clipped from 1:1 mixture of the two corresponding substrates. Both of the SNP sites can be precisely genotyped by one-run of mass analysis. The analysis is further facilitated by differentiating the length between the two acridines in the tandem-activators and preparing RNAs of different length, since overlapping of mass signals can be minimized.

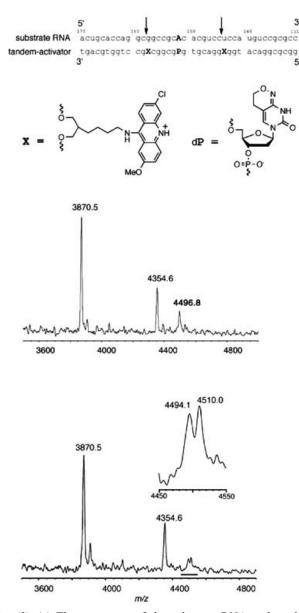


Fig. (3). (a) The structures of the substrate RNA and tandemactivator. (b) MALDI-TOF mass spectrum of the cleavage products of homozygous sample. (c) The spectra of the cleavage products of heterozygous sample. The assignment of each peak is described in the text.

In most of the previous methods for SNP genotyping, the substrate is DNA, which can be easily obtained by amplification of human genomes with PCR. However, DNA normally takes duplex form, which is hard to be analyzed. Thus, the PCR products (double-stranded DNA) must be converted to single-stranded one before the genotyping assay. In the present method, the substrate is RNA, and thus this procedure is unnecessary. Sufficient amount of singlestranded RNA can be obtained by transcription with RNA polymerase from duplex DNA, which is prepared by PCR with the use of promoter-appended primers in minimized reaction cycles.

3. SNP GENOTYPING WITH THE USE OF ARTIFICIAL RESTRICTION-ENZYME SYSTEMS

3.1 Strict Recognition of a Mismatch by the Combination of PNA and Nuclease S1

The second method is based on our new finding that the combination of peptide nucleic acid (PNA) [8] and singlestranded DNA specific nucleases [9] strictly recognize onebase alteration in DNA with remarkable specificity (Fig. 4). Peptide nucleic acid (PNA: Fig. 5), elegantly developed by Nielsen, is one of the most widely used DNA analogs. We found that DNA is hardly digested by the enzyme when it forms a duplex with completely complementary PNA. However, even one-base mismatch in DNA to PNA makes this DNA susceptible to the enzymatic digestion. In place of nuclease S1, nuclease P1 and Mung Bean nuclease can be also used.

The present molecular recognition of one-base alteration in DNA sequence by PNA takes advantage of notable structural and physicochemical differences between fully matching-site and mismatching-site in DNA/PNA duplexes. Dictated by this difference, single-stranded DNA specific enzymes selectively choose these mismatching-sites and hydrolyze the DNA there. This recognition is so precise and strict that the DNA is efficiently cleaved even when there exists only one mismatch between DNA and PNA. In contrast, fully-matching DNA/PNA duplex is sufficiently resistant to the enzyme.

3.2 Visual Detection of SNPs

These strict discriminations between full-match and mismatch by the PNA/nuclease S1 combination can be

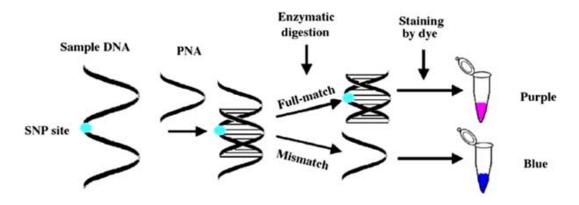


Fig. (4). Flowchart of SNP detection by PNA/nuclease/dye system

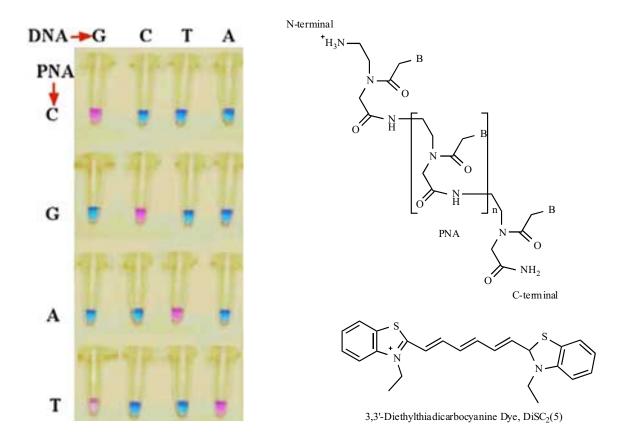


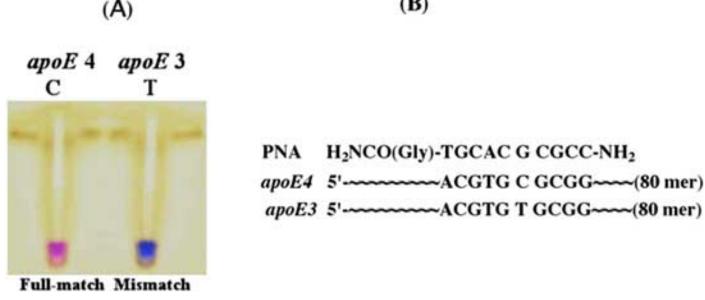
Fig. (5). Visual detection of mismatches by using $DiSC_2(5)$. The DNA/PNA duplexes were treated by nuclease S1 and then the mixtures were stained with $DiSC_2(5)$. Among the mismatches, only the T(PNA)-G(DNA) pair (the left in the bottom) took a pale purple color. However, this sample could be clearly differentiated from the matching ones by visible/UV spectroscopy (see ref. 4 for the detail).

easily and clearly visualized to naked eyes, by adding 3,3'-diethylthiadicarbocyanine DiSC₂(5) (the structure is presented in Fig. (5)) to the enzymatic digests. This dye changes its color from blue to purple when it is bound to DNA/PNA duplexes. [10] Thus, the enzymatic digest of completely complementary DNA/PNA duplex provides purple color. On the other hand, DNA having a mismatch to the PNA provides a blue color.

Typical examples are presented in Fig. (5). As shown in the top-left, the solution of enzymatic digest of DNA^(match) containing this dye takes purple color. The 10-mer portion of this DNA is complementary with the PNA, and thus is sufficiently protected from the enzymatic digestion. The DNA/PNA duplexes left in this solution bind the $DiSC_2(5)$ dye, giving rise to the purple color. On the other hand, $DNA^{(C-C)}$, $DNA^{(C-T)}$, and $DNA^{(C-A)}$ have a mismatch to the PNA, and thus are easily digested down to mononucleotides by the nuclease. There are no DNA/PNA duplexes left in the reaction mixtures, and therefore all of the solutions remain blue (the three tubes in the top row). Other mismatches also provide blue solutions and are clearly discriminated by naked eyes from the fully-matching pairs. Thus, mismatching DNAs are completely discriminated by naked eyes from fully-matching one. In Fig. (5), the T (in PNA)-G (in DNA) mismatch (the bottom left) shows a palepink color, probably because this mismatch forms a rather stable non-Watson-Crick base pair. According to reversedphase HPLC, the enzymatic digestion is not so efficient in this mixture and some amount of DNA/PNA duplex still remains in the final mixture. However, this T-G mismatch is explicitly distinguishable from T-A fully-matching pair by visible-absorption spectroscopy. Alternatively, the color difference can be greatly magnified to allow naked-eye detection simply by increasing the concentration of the dye (data not presented). Advantageously, contaminating DNA/DNA duplexes, if any, cause no significant problem, since this dye remains blue even in the presence of these duplexes [10]. The blue \rightarrow purple change is specific to DNA/PNA duplexes, and thus precisely reflects the perfect matching of sample DNA to the PNA probe.

3.3 Application to Visual Detection of SNPs in Human Genomic DNA

The method described above has been applied to the detection of SNPs in human *apoE* gene. The 80 mer DNA between the position 120 and 199 (containing the codon 112 SNP (TGC \rightarrow CGC) in *apoE* gene) is used. The SNPs site is the nucleotide-152, which is T in *apoE3* but C in *apoE4*. After amplification by asymmetric PCR, the single-stranded DNA containing this SNP site is treated at 20 °C for 30 min with nuclease S1 in the presence of PNA (H₂NCO(Gly)-



(**B**)

Fig. (6). (A) Detection of SNPs of human apoE gene by the PNA/nuclease S1/DiSC₂(5) system. As shown in (B), PNA is complementary with the part of *apoE*4 but has a G-T mismatch to *apoE*3.

TGCAC G CGCC-NH₂), which is complementary to the sequence in *apoE*4, and then the reaction mixture is stained with $DiSC_2(5)$. As shown in Fig. (6A), the solution from the allele of *apoE4* (completely complementary with this PNA) is purple, whereas the solution from the allele of apoE3 (containing a G-T mismatch) is blue. In other series of experiments, these two DNAs are treated with nuclease S1 in the presence of PNA, which is complementary with *apoE3*. Exactly as expected, the solution from the allele of *apoE3* is purple, and the solution from the allele of *apoE4* is blue. The kind of allele of the patient has been systematically pinned down.

3.4 PNA as Essential Cofactor for the Present Method

In order to achieve the present one-base differentiating protection and digestion, PNA must be used because (1) a mismatch in DNA/PNA duplex notably destabilizes the duplex, facilitating the enzymatic reaction at this site, and (2) PNA is resistant to the enzymatic digestions and functions as DNA-protecting agent throughout the reaction time. Accordingly, SNPs typing is easily and straightforwardly accomplishable by adding $DiSC_{2}(5)$ dye after digesting the DNA in the presence of appropriate PNA. One of the most important advantages of this method is its simplicity. Without any complicated equipment, SNPs are visually and conveniently detectable. When the solutions are purple, the DNA is completely matching with the PNA probe. Even with only one mismatch, however, blue solutions are obtained. The sensitivity of detection should be further improved by designing still more effective dye and/or by modifying the PNA appropriately. The present method should be also applicable to high through-put genotyping of multiple SNPs. In combination with array or beads technology, amplified DNA fragments in multiplex

reactions can be captured on specific PNA probes, and then subjected to the enzymatic digestion. Undigested DNA/PNA duplex will be detected with $DiSC_2(5)$ dye by absorption spectroscopy. The ratio of absorbance in the purple region to that in the blue region can be a reliable measure of mismatch. Such PNA probes on a solid surface are resistant to nuclease digestion and can be utilized repeatedly.

The length of PNA probe is crucially important for successful analysis. Hybridization of too short complementary sequences is thermodynamically unfavorable under common analytical conditions. On the other hand, too long PNA often easily aggregates and is also unfavorable. Preferably the length of PNA is 10-14 mer and the content of purine should be within the value of 25%~60% to avoid undesired aggregation. Self-complementary sequences in PNA is of course unacceptable. Although PNA forms a duplex with DNA in both anti-parallel and parallel orientations, the former is usually more stable. [11]

4. CONCLUSION

In this review, two methods for SNP genotyping, both of which have been recently developed by the authors' group, are presented. One of them involves selective scission of RNA at two-sites by artificial ribonuclease systems. By analyzing the resultant small RNA fragment by mass spectroscopy, precise genotyping is achievable. This method should be highly promising for multiplex analysis, since a number of RNA fragments of desired length can be prepared by the present method and simultaneously analyzed by mass spectroscopy. In the second method, mismatch between DNA sample and PNA probe is pinned down in terms of the stability against nuclease. The presence of oligonucleotide left (or its absence) in the enzymatic digests is visually scored by appropriate dye. Easiness and convenience are the striking advantages of this method.

ACKNOWLEDGEMENTS

This work was partially supported by the Biooriented Technology Research Advancement Institution. The support by a Grant-in-Aid for Scientific Research from the Ministry of Education, Science, Culture, Sport and Technology, Japan is also acknowledged.

REFERENCES

- (a) Syvänen, A.C. Nat. Rev. Genet. 2001, 2, 930-942. (b) Kirk, B.W.; Feinsod, M.; Favis, R.; Kliman, R.M.; Barany, F. Nucleic Acids Res. 2002, 30, 3295-3311. (c) Lyamichev, V.; Mast, A.L.; Hall, J.G.; Prudent, J.R.; Kaiser, M.W.; Takova, T.; Kwiatkowski, R.W.; Sander, T.J.; de Arruda, M.; Arco, D.A.; Neri, B.P.; Brow, M.A.D. Nat. Biotechnol. 1999, 17, 292-296. (d) Alderborn, A.; Kristofferson, A.; Hammerling, U. Genome Res. 2000, 10, 1249-1258.
- [2] Kuzuya, A.; Mizoguchi, R.; Morisawa, F.; Komiyama, M. Chem. Commun. 2003, 770-771.
- [3] (a) Kuzuya, A.; Mizoguchi, R.; Morisawa, F.; Machida, K.; Komiyama, M. J. Am. Chem. Soc. 2002, 124, 6887-6894. (b) Kuzuya, A.; Machida, K.; Mizoguchi, R.; Komiyama, M. Bioconj. Chem. 2002, 13, 365-369.
- [4] Komiyama, M.; Ye, S.; Liang, X.; Yamamoto, Y.; Tomita, T.; Zhou, J.-M.; Aburatani, H. J. Am. Chem. Soc. 2003, 125, 3758-3762.
- [5] Recent reviews: (a) Komiyama, M; Sumaoka, J; Kuzuya, A; Yamamoto, Y. *Methods Enzymol.* 2001, 341, 455-468. (b) Trawick, B.N.; Daniher, A.T.; Bashkin, J.K. *Chem. Rev.* 1998, 98, 939-960. (c) Oivanen, M.; Kuusela, S.; Lönnberg, H. *Chem. Rev.*

1998, *98*, 961-990. (d) Perreault, D.M.; Anslyn, E.V. Angew. Chem. Int. Ed. Engl. **1997**, *36*, 432-450.

- [6] Kuzuya, A.; Machida, K.; Komiyama, M. Tetrahedron Lett. 2002, 43, 8249-8252.
- [7] (a) Mahley, R.; Rall, S.C. Ann. Rev. Genomics Hum. Genet. 2000, *I*, 507-537. (b) Jacobsen, N.; Bentzen, J.; Meldgaard, M.; Jakobsen, M.H.; Fenger, M.; Kauppinen, S.; Skouv, J. Nucleic Acids Res. 2002, 30, e100. (c) Nickerson, D.A.; Taylor, S.L.; Fullerton, S.M.; Weiss, K.M.; Clark, A. G.; Stengård, J.H.; Salomaa, V.; Boerwinkle, E.; Sing, C.F. Genome Res. 2000, 10, 1532-1545. (d) Vanduijn, C.M.; Deknijff, P.; Cruts, M.; Wehnert, A.; Havekes, L.M.; Hofman, A.; Vanbroeckhoven, C. Nat. Genet. 1994, 7, 74-78.
- [8] (a) Nielsen, P.E.; Egholm, M.; Berg, R.H.; Buchardt, O. Science 1991, 254, 1497-1500. (b) Nielsen, P.E. Pure Appl. Chem. 1998, 70, 105-110. (c) Nielsen, P.E.; Haaima, G. Chem. Soc. Rev. 1997, 26, 73-78.
- [9] (a) Desai, N.A.; Shankar, V. FEMS Microbiol. Rev. 2003, 26, 457-491. (b) Romier, C.; Dominguez, R.; Lahm, A.; Dahl, O.; Suck, D. Proteins 1998, 32, 414-424. (c) Suck, D. Biopolymers 1997, 44, 405-421. (d) Chaudhry, M.A.; Weinfeld, M. Nucleic Acids Res. 1995, 23, 3805-3809. (e) Oleson, A.E.; Hoganson, E.D. Arch. Biochem. Biophys. 1981, 211, 478-484.
- [10] (a) Smith, J.O.; Olson, D.A.; Armitage, B.A. J. Am. Chem. Soc. 1999, 121, 2686-2695. (b) Datta, B.; Armitage, B.A. J. Am. Chem. Soc. 2001, 123, 9612-9619. (c) Wilhelmsson, L.M.; Nordén, B.; Mukherjee, K.; Dulay, M.T.; Zare, R.N. Nucleic Acids Res. 2002, 30, e3.
- [11] (a) Tomac, S.; Sarkar, M.; Ratilainen, T.; Wittung, P.; Nielsen, P.E.; Nordén, B.; Graslund, A. J. Am. Chem. Soc. 1996, 118, 5544-5552. (b) Eriksson, M.; Nielsen, P.E. Q. Rev. Biophys. 1996, 29, 369-394. (c) Egholm, M.; Buchardt, O.; Christensen, L.; Behrens, C.; Freier, S.M.; Driver, D.A.; Berg, R.H.; Kim, S.K.; Nordén, B.; Nielsen, P.E. Nature 1993, 365, 566-568. (d) Haaima, G.; Lohse, A.; Buchardt, O.; Nielsen, P.E. Angew Chem. 1996, 35, 1939-1942. (e) Rose, D.J. Anal. Chem. 1993, 65, 3545-3549.