Award Accounts

The Chemical Society of Japan Award for Creative Work for 2007

Recognition of Mismatched Base Pairs in DNA

Kazuhiko Nakatani

The Institute of Scientific and Industrial Research, Osaka University, 8-1 Mihogaoka, Ibaraki, Osaka 567-0047

Received March 17, 2009; E-mail: nakatani@sanken.osaka-u.ac.jp

Mismatch binding ligand (MBL) selectively binds to mismatched base pairs with discrimination of the nucleotide bases composing the mismatch. This review focuses on the inside story regarding the molecular design of MBL and its function, and potential application in genomic science.

1. Introduction

DNA is one of the most beautiful molecular architectures. The right-handed helix of typical B-form duplex attracted much attention of scientists to explore the way to inherit the genetic information between generations, the structure–function relations, and the way to use DNA as building blocks in nano science. Synthetic organic chemists have long looked for molecules that can recognize the base sequence and eventually modulate the flow of genetic information.

Minor groove binding and intercalation between base pairs are the major mode of DNA recognition by small molecules. Among numerous studies on DNA recognition, pyrrole–imidazole polyamide developed on the basis of the mode of DNA binding of natural product distamycine and almost twenty years of studies on synthetic derivatives are now available as the molecular tool for the sequence-specific recognition of DNA, the modulation of gene expression, and the assembly of nanoscale DNA structures.[1–3]

In contrast to these studies that focused on the recognition of DNA without structural deficits, there was little study that targeted the recognition of imperfect DNA structures. DNA is an organic compound consisting of atoms of C, H, N, O, and P. Therefore, DNA must follow the fate of organic molecules, thus, DNA is not resistant to oxidation and chemical reactions. DNA could be easily oxidized especially at the guanine base.[4] Guanines and adenines also react with a variety of electrophiles to give DNA with modified or damaged bases,[5] which eventually eliminates the modified base moiety to produce abasic sites. Upon UV exposure, a continuous T sequence could be transformed into a thymine dimer,[6] which is known as a cause of skin cancer. The cytosine base undergoes deamination under physiological conditions to the uracil.[7] The C to U deamination will lead to the genetic mutation of C to T transversion. Our genome consisting of 3 billon base pairs is synthesized by DNA polymerase. In addition to these chemical reactions of nucleotide bases, error in enzymatic reactions causes a deficit in the DNA structure. DNA polymerase must accurately polymerize nucleoside triphosphates according to the base sequence of the template DNA. Incorporation of mismatched nucleosides is however, inevitable no matter how high the fidelity of DNA polymerase reaction is.[8] Chemical degradations of nucleotide bases and errors in polymerase reactions result in the formation of imperfect double-stranded DNA having damaged nucleotide bases, abasic sites, bulged bases, and mismatched base pairs (Figure 1). Because these non-canonical structures accelerate the mutation of genetic information, nature prepares enzymatic repair systems to recognize and get rid of the imperfect DNA structures from the genome, and restore the correct genetic information.[9]

We were particularly interested in the mode of recognition of mismatched base pairs by repair enzymes. Uracil DNA glycosylase (UDG) recognizes the G–U mismatched base pair in duplex DNA and removes the uracil base from the DNA strand.[7] The co-crystal of UDG and DNA duplex is quite informative how the enzyme discriminates the G–U mismatched base pair from Watson–Click base pairs and removes the uracil from DNA (Figure 2). UDG intrudes the amino acid side chain in front of the guanine and lets the uracil base flip

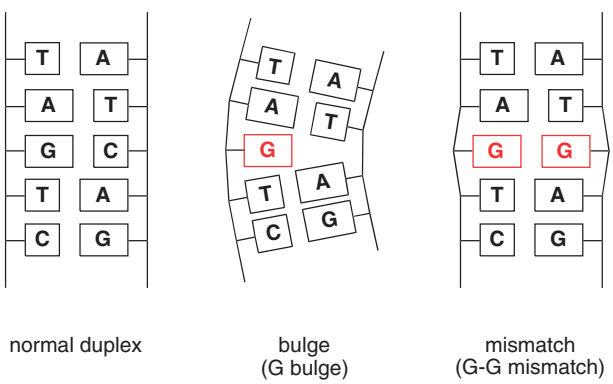


Figure 1. Illustrative representations of normal duplex, bulge, and mismatch.

Published on the web September 9, 2009; doi:10.1246/bcsj.82.1055

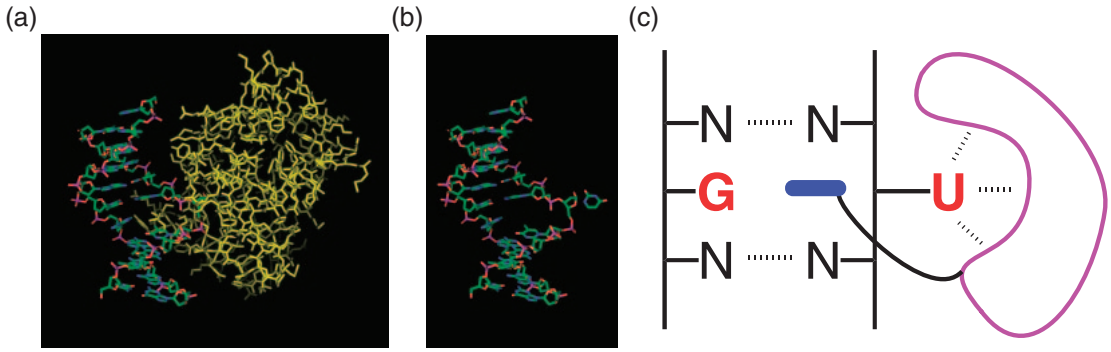


Figure 2. Binding mode of uracil DNA glycosylase. (a) X-ray structure of co-crystal, (b) DNA part of X-ray structure, and (c) cartoon of the mode of UDG binding.

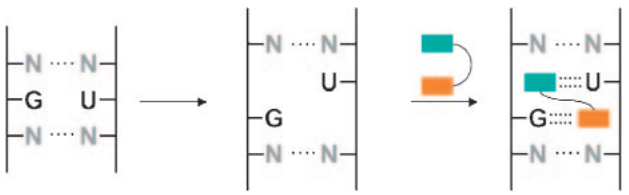


Figure 3. A proposed motif for the mismatch recognition. Colored rectangular objects indicate recognition elements.

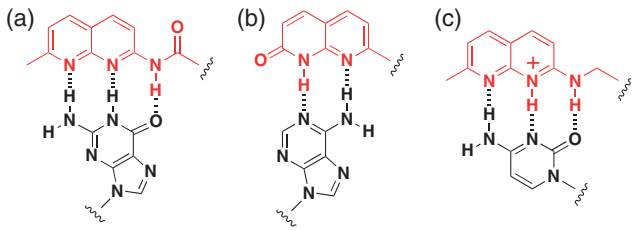


Figure 4. Recognition elements used for MBLs. (a) *N*-Acyl-2-amino-7-methyl-1,8-naphthyridine for G, (b) 7-methyl-2-oxo-1,8-naphthyridine (azaquinolone) for A, and (c) protonated 2-amino-7-methyl-1,8-naphthyridine for C.

out from the base π-stack.[10] The flipped out uracil base is captured by the recognition pocket of UDG through hydrogen bonds and other interactions. Upon flipping out, the glycosidic bond between uracil N1 and C1′ of the deoxyribose becomes accessible from the backside by the enzyme side chain to accelerate the hydrolysis of the uracil. The base flipping mechanism for the recognition of the mismatched base pairs by the repair enzymes is quite ingenious. We asked ourselves a question how we could achieve mismatch recognition with small organic molecules. Mismatch recognition by the base flipping mechanism is suitable for large proteins but not for small organic molecules.

When we started the research project aiming at molecular design for mismatch recognition, a few reports described the interaction of small molecules with unpaired nucleotide bases. Natural products neocarzinostatin chromophore and bleomycin were known to bind to the bulged sites in duplex DNA.[11] Rhodium intercalators were also demonstrated to bind to mismatched base pairs.[12] However, there were no precedents of reports describing the recognition of bulged sites and mismatched base pairs with base discrimination.

2. Conceptual Model for Mismatch Recognition

We have proposed a novel motif for mismatch recognition by assuming a hypothetical structure of mismatched base pairs consisting of two successive bulged bases (Figure 3).[13–15] This hypothetical structure would be stabilized by a molecular ligand that occupies the space and forms the hydrogen bonds to the mismatched nucleotide bases. Furthermore, the pseudo base pairs could be further stabilized by stacking interaction with the neighboring base pairs. The key elements of mismatch recognition are planar and flat heterocyclic molecules having hydrogen-bonding capability to the nucleotide bases. We call these heterocycles recognition elements. It has been reported that *N*-acyl-2-amino-1,8-naphthyridine derivatives bind to guanine by forming three hydrogen bonds (Figure 4).[16–18] In pure chloroform, the association constant is almost comparable to that for the G–C base pair formation. We anticipated that a naphthyridine derivative that is soluble in water and has some positive charges might bind to the unpaired guanine in duplex DNA, because the inner space of DNA is anticipated to be rather more hydrophobic than the hydrophilic environment of an outer space of the duplex. The proposed concept of mismatch recognition is quite flexible in terms of molecular design. According to the concept, molecules that bind to any one of eight mismatched base pairs could be designed by simply choosing an appropriate recognition element for the nucleotide bases in the mismatch. Accordingly, we have used 8-azaquinolone as the recognition element for the adenine,[19,20] and protonated form of 2-amino-1,8-naphthyridine as the recognition element for the cytosine.[21] The mismatch binding ligands (MBLs) we have so far discovered are shown in Figure 5. In this article, we describe the inside story of MBLs about the discovery, identifying the function, and the application in genomic science.[22]

3. Recognition of Mismatched Base Pairs

3.1 G–G Mismatch Recognition. The first compound we designed was the dimeric form of *N*-acyl-2-amino-1,8-naphthyridine which we named naphthyridine dimer ND.[13,14] ND consisted of two naphthyridine heterocycles and a linker connecting them. The linker contained a secondary amino group, which functions as a positively charged site to gain water solubility and provides attractive electrostatic interaction to negatively charged DNA, and also as a site of tethering for the surface immobilization of the ligand. However, the most important role of the linker is to offer the appropriate

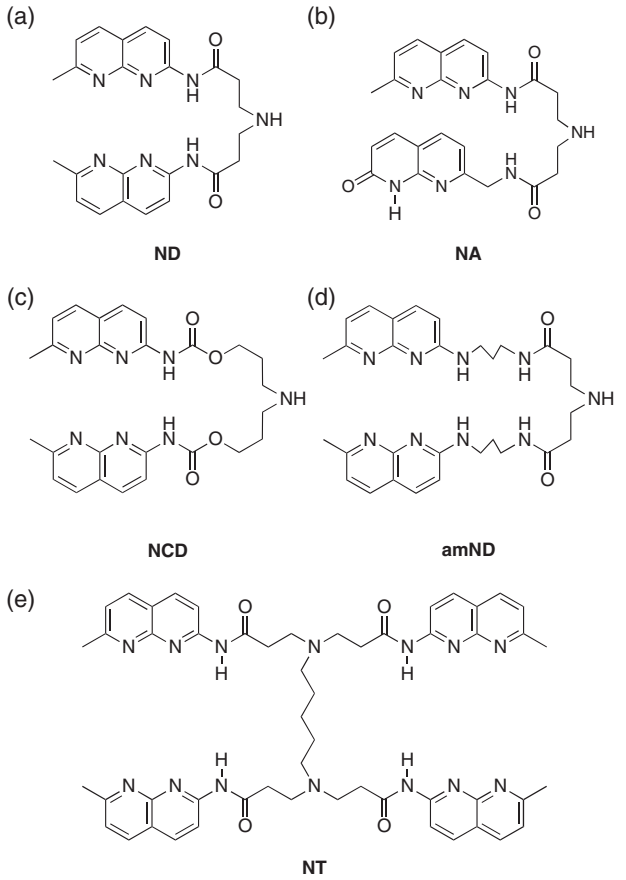


Figure 5. Structures of MBLs. (a) naphthyridine dimer (ND), (b) naphthyridine–azaquinolone (NA), (c) naphthyridine carbamate dimer (NCD), (d) aminonaphthyridine dimer (amND), and (e) naphthyridine tetramer (NT).

conformational restriction to the dynamic motion of two heterocycles.[23] The concept of the linker design was common for all MBLs.

ND was synthesized by a coupling of 2-amino-7-methyl-1,8-naphthyridine and a linker moiety.[13,15,22] The selective binding of ND to the G–G mismatch, among seven other mismatches, has been verified by examining the thermal stability of the mismatch-containing duplexes in the presence of ND. In general, ligand binding to DNA increases the thermal stability of the duplex, making the temperature required for denaturation of a duplex to a single strand higher than that in the absence of the ligand. The melting temperature (T_m) necessary for 50% denaturation of the G–G mismatch-containing duplex increased by 16.4 °C in the presence of ND, whereas only small increases were observed for duplexes containing other mismatches, as well as for fully complementary duplexes (Table 1).[13] The T_m of the G–G mismatch-containing duplex does not increase in the presence of mono-naphthyridine, which can stabilize DNA containing a G-bulge,[15] suggesting that the cooperative and simultaneous binding of two naphthyridines of ND to both G bases is extremely important for stabilization of the complex. The dissociation constant (K_d) for the complex of ND and the G–G mismatch in the 5′CGG3′/5′CGG3′ sequence was determined to be 53 nM, by quantitative DNase I footprint titration experiments (Figure 6).[24] Under these experimental conditions, the binding of ND to the G–G mismatch is about 350-fold more efficient than is binding to the G–A (K_d = 19 μM) and G–T (K_d = 21 μM) mismatches in 5′CGG3′/5′CAG3′ and 5′CGG3′/5′CTG3′ sequence, respectively. These data are fully consistent with those of duplex stabilization by ND, obtained by T_m measurements.

Table 1. T_m of the Mismatch-Containing Duplex in the Presence of ND[a)]

5′-CTAAC**X**GAATG-3′
3′-GATTG**Y**CTTAC-5′

X–Y	$T_{m(-)}$	$T_{m(+)}$	ΔT_m
G–G	25.8	42.2	16.4
G–A	27.0	29.3	2.3
G–T	28.7	30.8	2.1
A–A	20.0	20.5	0.5
A–C	18.0	19.8	1.8
T–T	26.8	26.8	0.0
T–C	20.5	21.7	1.2
C–C	20.8	23.0	2.2
G–C	42.5	42.2	−0.3
A–T	37.5	36.8	−0.7

a) The UV-melting curve was measured for a duplex at a total base concentration of 100 μM in 10 mM sodium cacodylate buffer (pH 7.0) containing 0.1 M NaCl. Melting temperature in the absence ($T_{m(-)}$) or presence ($T_{m(+)}$) of ND (9.1 μM).

The energetic parameters and stoichiometry of the interaction between ND and 11-mer duplexes d(CTA ACG GAA TG)/d(CAT TCG GTT AG) containing a G–G mismatch were analyzed by isothermal titration calorimetry (ITC).[25] An apparent association constant (K_a) for the binding of ND to the G–G mismatch was $0.9 \times 10^7\,M^{-1}$, that is consistent with K_a of $1.9 \times 10^7\,M^{-1}$ (K_d of 53 nM) obtained by DNase I footprint titration within an experimental variability. The stoichiometry (n) for the binding was 1.2, suggesting a one to one binding. The free energy change at 25 °C by the formation of the complex was -9.5 ± 0.3 kcal mol^{-1} (1 kcal mol^{-1} = 4.184 kJ mol^{-1}), which consisted of an enthalpy gain of -36.5 ± 0.4 kcal mol^{-1} and an entropy loss of 27.0 ± 0.7 kcal mol^{-1}. These data showed an enthalpy-driven binding of ND to a G–G mismatch. While we have concluded the one to one binding based on these experimental data, we later found by mass spectrometry of the complexes that ND binding to the G–G mismatch in fact involved both one to one and two to one binding. The extraordinary and exclusively two to one binding will be discussed later.

CD spectra of the G–G mismatch were measured in the presence and absence of ND (Figure 7). While CD spectra of the G–G mismatch is quite similar to that of fully matched duplex, addition of two equivalent moles of ND induced a dramatic change of the spectra.[14] The positive maximum at 270 nm increased the intensity by twofold and negative bands at 250 nm shifted to 230 nm in the complex. Intense induced CD was also observed in the region from 300 to 360 nm, showing that naphthyridine rings are located within a chiral environment produced by a DNA double helix. These CD experiments clearly showed that binding of ND to a G–G mismatch accompanied a significant conformational change.

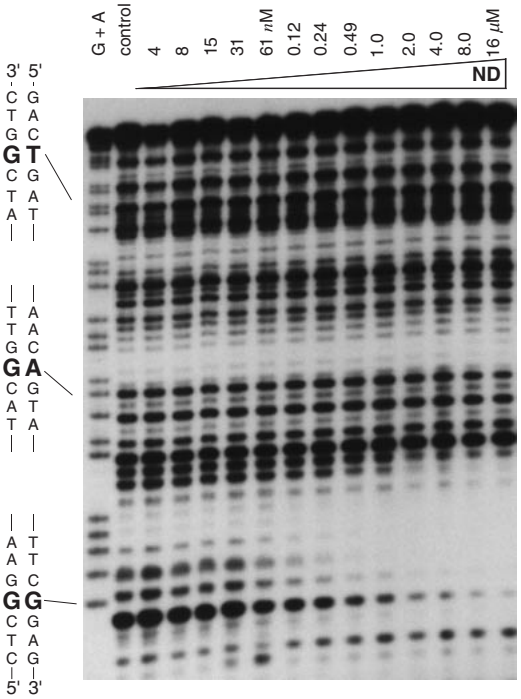


Figure 6. DNase I footprinting analysis of the complex of ND and 5′-^{32}P-end-labeled 64-mer DNA duplex containing G–G, G–A, and G–T mismatched sites. DNA duplex (4 nM, 1×10^3 cpm) of 5′-^{32}P-end-labeled 64-mer d(GTT ACA GAA TCT CGG AAG CCT AAT ACG GTT GAC ATT CAC GGT TTC ACT GAT CGG TCT GAG TAG C) and a complementary 64-mer d(GCT ACT CAG ACT GAT CAG TGA AAC CGT GAA TGT CAA CAG TAT TAG GCT TCG GAG ATT CTG TAA C) was incubated with various concentrations of ND (4 nM–16 μM) in a Tris-HCl buffer (10 mM, pH 7.6) containing NaCl (100 mM) and $MgCl_2$ (5 mM) for 12 h at 4 °C. DNase I (0.2 units) was added to the mixture and incubated for 8 min at 25 °C. Recovered DNA by ethanol precipitation was analyzed by electrophoresis on a denaturing sequencing gel containing 12% polyacrylamide and 7 M urea. Lane 1, Maxam-Gilbert A + G sequencing reaction; lanes 2–15, 0, 4, 8, 15, 31, 61 nM, 0.12, 0.24, 0.49, 1.0, 2.0, 4.0, 8.0, and 16 μM of ND, respectively. Partial sequences of DNA with an emphasis of the mismatched sites shown on the left side of the gel.

While two naphthyridine heterocycles were connected by an amide linkage, the amide bond was rather easily cleaved off especially by the intramolecular attack of the secondary amino group. This was most likely due to the strong electron-withdrawing nature of the 1,8-naphthyridine. To improve the thermal stability of ND and also expand the planer surface, we synthesized NCD in which two naphthyridine heterocycles were connected by a carbamate linkage.[26–28] Under the alkaline conditions of 50 mM sodium hydroxide at room temperature, only 29% of ND remained after 2 h incubation, whereas 86% of NCD still remained. After incubation for 4 h, the remaining ND and NCD was 7% and 79%, respectively. The binding of NCD to mismatches was compared to that of ND by the ΔT_m. NCD showed an improved binding and selectivity to the G–G mismatch (Table 2). We later discuss an extraordinary NCD binding to the CGG/CGG triad.[28]

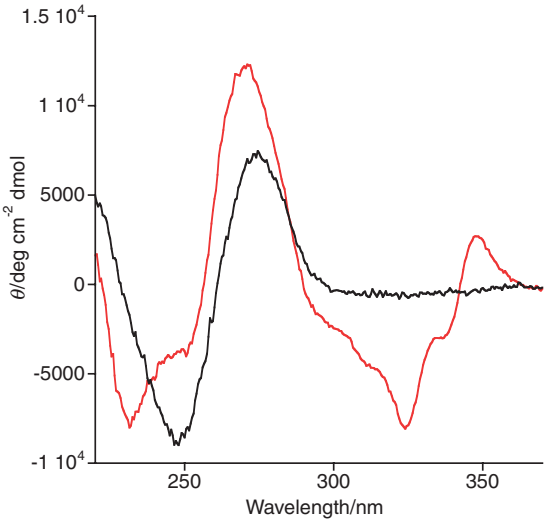


Figure 7. CD spectra of DNA duplexes d(CTA ACG GAA TG)/d(CAT TCG GTT AG) containing a G–G mismatch in the absence (black) and presence (red) of ND (9.1 μM). CD spectra of duplexes (100 μM base concentration) were measured in 10 mM sodium cacodylate buffer (pH 7.0) and 100 mM NaCl at 25 °C.

Table 2. ΔT_m of Mismatch-Containing Duplexes in the Presence of NCD[a)]

5′-CTAAC**X**GAATG-3′
3′-GATTG**Y**CTTAC-5′

X–Y	$T_{m(-)}$	ΔT_m	
		NCD	ND
A–A	17.8 (1.4)	1.5 (0.2)	−0.8 (1.1)
A–C	16.1 (0.8)	4.1 (0.2)	2.1 (0.2)
C–C	18.2 (0.2)	6.1 (0.3)	6.7 (0.6)
G–A	25.7 (0.2)	7.0 (0.2)	8.6 (1.3)
G–G	25.6 (1.3)	29.1 (0.2)	23.7 (1.2)
G–T	28.3 (0.2)	1.2 (0.8)	10.2 (1.3)
T–C	18.6 (0.2)	3.7 (0.3)	5.5 (0.7)
T–T	25.1 (0.2)	0.7 (0.7)	0.6 (1.2)
A–T	34.3 (0.3)	0.0 (0.3)	−1.5 (0.3)
G–C	40.3 (0.0)	−2.0 (0.4)	2.0 (0.9)

a) The UV-melting curve was measured for a duplex at a total base concentration of 100 μM in a 10 mM sodium cacodylate buffer (pH 7.0) containing 0.1 M NaCl. All measurements were taken three times, and standard deviations are shown in parenthesis. ΔT_m is calculated as a difference of T_m in the presence and absence of drugs (100 μM), respectively.

3.2 G–A Mismatch Recognition. Taking into account the structure of ND as a clue for the molecular design of ligands binding to G–A mismatches, we have discovered naphthyridine–azaquinolone hybrids (NA) where one naphthyridine chromophore in ND is replaced by an 8-azaquinolone chromophore having complementary hydrogen-bonding surfaces to that of adenine (Figure 5b).[19] The structure–activity relationship for the binding of NA hybrid to G–A mismatches was studied with hybrid molecules consisting of quinoline and aminoquinoline as recognition elements (Figure 8).

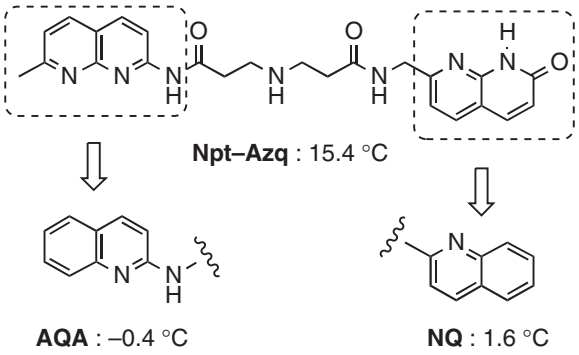


Figure 8. Illustrative representation of the effect of structural modification of chromophores in NA on the increased melting temperature (ΔT_m) of the G–A mismatch.

Table 3. ΔT_m Values for the Duplexes Containing G–Y Mismatches in the Absence and Presence of Drugs[a)]

5′-CTAAC**G**GAATG-3′
3′-GATTG**Y**CTTAC-5′

G–Y	$T_{m(-)}$	ΔT_m			
		NA	ND	NQ	AQA
G–A	25.8	15.4	12.8	1.6	−0.4
G–G	25.1	10.6	26.2	4.5	0.1
G–C	38.6	2.2	4.3	2.0	0.7

a) T_m for the duplex (5.0 μM each strand) in the absence and presence of ligand (200 μM) was measured in 10 mM sodium cacodylate buffer (pH 7.0) containing 100 mM NaCl.

The binding of NA and other ligands to G–A and G–G mismatches was examined by measuring T_m of mismatch-containing duplexes. Under the conditions, NA increased the T_m of the G–A mismatch by 15.4 °C, whereas modest ΔT_m of 10.6 °C was observed for the G–G mismatch (Table 3). NA did not stabilize the full match duplex. ND stabilized G–A and G–G mismatches by 12.8 and 26.2 °C, respectively. The substitution of one recognition element markedly changed the spectrum for the mismatch recognition. Two reference compounds NQ and AQA stabilized neither the G–A mismatch nor the G–G mismatch. These remarkable effects of the substitution of the recognition elements in NA and ND on the stabilization of the G–A mismatch suggests the significance of the hydrogen bonding of 8-azaquinolone to adenine and acylated-2-amino-1,8-naphthyridine to guanine (Figure 4). We later found that NA bound to the A–A mismatch with extremely high sequence specificity. We will discuss the extraordinary NA binding to the CAG/CAG triad in a following chapter.[20]

3.3 C–C Mismatch Recognition. The recognition of cytosine requires a hydrogen-bonding donor in the middle of an alignment of donor–donor–acceptor groups that is complementary to the alignment of acceptor–acceptor–donor groups of cytosine. After many trials for recognition elements, we eventually found that *N*-alkyl-2-amino-1,8-naphthyridine strongly and selectively bound to the cytosine–cytosine mismatch.[21] The ligand we have developed for the recognition of the C–C mismatch was the dimeric form of aminonaphthyridine (amND).

Table 4. ΔT_m Values for the 11-mer Duplexes Containing a Mismatch[a)]

5′-CTAAG**X**CAATG-3′
3′-GATTC**Y**GTTAC-5′

X–Y	T_m	ΔT_m
C–C	18.0 (0.4)	18.0 (0.4)
C–T	22.3 (0.5)	11.3 (0.3)
T–T	23.0 (0.8)	7.2 (0.6)
C–A	27.6 (0.4)	4.0 (0.1)
A–A	24.2 (0.4)	1.8 (0.8)
G–T	29.8 (0.2)	0.5 (0.7)
G–A	32.6 (0.6)	0.3 (0.8)
G–G	36.1 (0.5)	−0.1 (1.0)
A–T	37.0 (0.3)	−0.1 (0.1)
C–G	43.1 (0.2)	−0.3 (0.5)

a) T_ms of duplexes (4.5 μM) were measured in 10 mM sodium phosphate buffer (pH 7.0) containing 100 mM NaCl. All measurements were taken three times, and standard deviations are shown in parenthesis. The concentration of amND was 100 μM.

The remarkable increase of T_m by 18.0 °C was observed for the C–C mismatch in the presence of amND, whereas the T_ms of fully matched duplexes were not increased at all under the same conditions (Table 4). An increase of T_m by 11.3, 7.2, and 4.0 °C were observed for C–T, T–T, and C–A mismatches, respectively. Other mismatches including A–A, G–T, and G–A, showed little increase of their T_ms. These results indicated that amND strongly stabilized the C–C mismatch and the mismatches containing a cytosine and thymine with a reduced efficiency, but not at all for mismatches consisting of only purine nucleotide bases.

Effects of sequences flanking a C–C mismatch on amND binding were examined with 13-mer duplex 5′-d(GCTAA **XCZ** AATGA)-3′/3′-d(CGATT **YCW** TTACT)-5′ containing a C–C mismatch in the sequence of 5′-**XCZ**-3′/3′-**YCW**-5′, where **X–Y** and **Z–W** were any combination of Watson–Crick base pairs. The ΔT_m values obtained for all sequences were over 10 °C, indicating that amND could stabilize the C–C mismatches regardless of the flanking sequences (Table 5). Rather low ΔT_ms of 10.4 and 11.2 °C recorded for **cCg**/**gCc** and **cCa**/**gCt** were due to the high T_m values of the duplex.

Since both ND and amND have the same hydrogen-bonding surface complementary to guanine but not to cytosine, the striking differences in base selectivity between two ligands needs rational explanation. The pH dependency of the UV absorbance of 2-amino-1,8-naphthyridine derivative revealed that the pK_a of the protonated form is about 6.8, showing that about 40% of the 2-aminonaphthyridine chromophore in amND in a free state would be protonated at pH 7.0 (Figure 9). Protonation of the chromophore at the N^1 position modulates from the hydrogen-bonding surface complementary to guanine to that to cytosine (Figure 10). It is reasonable to estimate that the protonation of N^1 in a bound state of 2-aminonaphthyridine to cytosine at neutral pH would be more facile than the protonation in a free state. While hydrogen bonding between non-protonated aminonaphthyridine and cytosine is feasible though two hydrogen bonds, energy gain by the hydrogen

Table 5. ΔT_m Values for the 13-mer Duplexes Containing a C–C Mismatch in Different Flanking Sequences[a)]

5′-GCTAA**XCZ**AATGA-3′
3′-CGATT**YCW**TTACT-5′

XCZ	YCW	$T_{m(-)}$	ΔT_m
GCC	CCG	25.4 (0.3)	17.9 (0.6)
GCT	CCA	23.9 (0.3)	15.7 (0.2)
CCC	GCG	22.3 (0.3)	15.3 (0.7)
GCA	CCT	26.2 (0.4)	15.0 (0.5)
ACT	TCA	20.7 (0.4)	14.2 (0.4)
ACA	TCT	23.2 (0.3)	13.5 (0.1)
CCT	GCA	25.3 (0.1)	13.2 (0.2)
TCA	ACT	24.8 (0.5)	12.2 (0.7)
CCA	GCT	29.0 (0.1)	11.2 (0.3)
CCG	GCC	32.6 (0.2)	10.4 (0.2)

a) Conditions of T_m measurements were the same as those in Table 4.

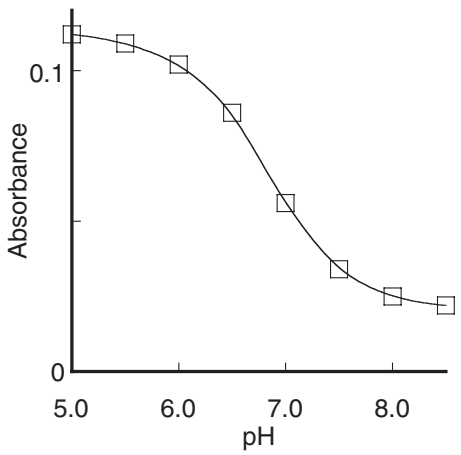


Figure 9. The pH dependency of a UV absorbance of a naphthyridine derivative at 370 nm. UV absorbance was measured in 10 mM sodium phosphate buffer in the presence of the ligand (10 μM) at various pH conditions.

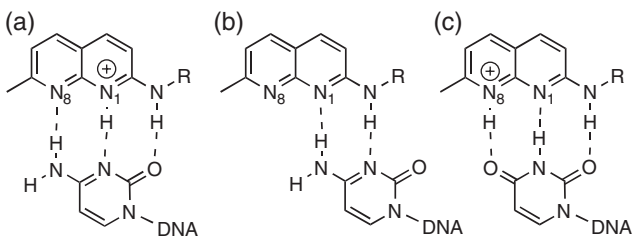


Figure 10. Proposed models of base-pairing between (a) cytosine and N^1-protonated 2-aminonaphthyridine, (b) cytosine and 2-aminonaphthyridine, and (c) thymine and N^8-protonated 2-aminonaphthyridine.

bonding calculated at the B3LYP/6-31G** level was 6.9 kcal mol^{-1} smaller than that between 1-methylcytosine and N^1-protonated 2-methylamino-7-methylnaphthyridine. On the basis of these arguments, the remarkably high selectivity of amND to the C–C mismatch is most likely due to a facile protonation of the chromophore under neutral pH.

The presence of a carbonyl group next to the exocyclic amino group in 2-aminonaphthyridine is responsible for the remarkable difference in binding selectivity to the mismatch. The strength of a hydrogen bond is related to hydrogen-bonding acidity of the donor group involved. Beijer et al. have reported that the association constant of the complex of 2,6-diaminopyridine with *N*-propylthymine increases approximately 10-fold upon acylation of the amino groups.[29] It is highly likely that an acylation of the exocyclic amino group of 2-aminonaphthyridine significantly improve the stability of the hydrogen-bonding pair to a guanine. Since the acylated aminonaphthyridine is less susceptible to the protonation at N^1 and N^8, the dimer of acylated naphthyridine showed negligible binding to the C–C mismatch.

Protonation of 2-amino-1,8-naphthyridine is feasible both at N^1- and N^8-positions. While N^1-protonation produces the hydrogen-bonding surface that is complementary to that of cytosine as discussed above, N^8-protonation produces a surface complementary to that of thymine (Figure 10c). This is most likely the molecular basis for the binding of amND to the C–T mismatch. The decreased binding to the C–T mismatch compared to the binding to the C–C mismatch was rationalized by an unfavorable secondary interaction in the hydrogen-bonded complex with alternating donor (D) and acceptor (A) groups. Studies on secondary interaction in triply hydrogen-bonded complex by Jorgensen et al. predicted that the base pair of ADA/DAD (e.g., T-N^8-protonated 2-aminonaphthyridine) is 11.3 kcal mol^{-1} less stable than that of DDA/AAD (e.g., C-N^1-protonated 2-aminonaphthyridine).[30,31]

The concept of spontaneous protonation to produce a hydrogen-bonding surface complementary to that of cytosine was quite useful. We later succeeded in developing a ligand selectively binding to a cytosine bulge.[32] The fluorescent ligand was particularly useful for discrimination of a single nucleotide difference of genome.[33]

4. Recognition of Repeat Sequences

There are a number of repeat sequences in our genome, where short and unique sequences are repeated many times. The number of repeats is often a determinant of its function. During the course of our studies on mismatched base pairs in double-stranded DNA, we faced a limitation in applying the mismatch binding ligands in biologically relevant systems. Mismatched base pairs are potential causatives of genetic mutation, and therefore are removed by repair systems as we discussed earlier. However, we reached an idea that a hairpin secondary structure of a single-stranded region of guanine rich human telomeric sequence may produce many G–G mismatches.[34–37] Later, we realized that hairpin structures of trinucleotide repeat sequences (TRS) consisted of a number of mismatched base pairs, and play in fact important roles in the expansion and contraction of TRS eventually leading to the onset of neurological disorders.[21,29] Ligands selectively binding to the particular repeat sequence would be useful as a molecular tool for the deeper understanding of the biological function and may eventually led to therapeutics for diseases related to the repeats.

4.1 Human Telomere Repeats. The telomeric sequence $d(TTAGGG)_n$ located at the 3′ end of human genomic DNA plays important roles in protecting chromosomal ends from

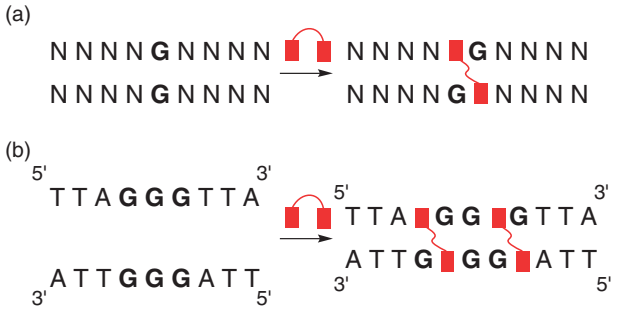


Figure 11. Illustrations of (a) a zigzag intercalation of ND into a duplex containing a single G–G mismatch and (b) a proposed binding of ND to a pair of a single-stranded overhang of a human telomeric sequence.

fusion, rearrangement, and translocation.[38,39] In normal cells, the entire length of the telomere gradually decreases as the replication of the genome is repeated,[40,41] due to the inability of DNA polymerase to replicate the very 3′ end of the template. In contrast, the enzyme telomerase is activated in the nuclei of cancer cells and maintains the length of the telomere for achieving immortality. Therefore, the ligands that bind to telomere and inhibit its elongation are expected to be potential anti-cancer drugs.[42,43] The single-stranded region of the d(TTAGGG)$_n$ repeat is known to form G-quadruplex structures in vitro.[44–46] A number of molecular ligands that bind to and strengthen the G-quadruplex structure have been investigated.[42,43] We anticipated that ND should act as molecular glue in the assembly of two telomeric sequences by binding strongly to G–G mismatches in the hypothetical duplex of a telomeric dimer (Figure 11).[34]

The DNA 22-mer d(AGGGTTAGGGTTAGGGTTAGGG) (telo22), consisting of four tandem repeats of a human telomeric sequence, forms an anti-parallel G-quadruplex[47] that shows typical CD spectra with positive and negative CDs at 290 and 265 nm, respectively.[48] Upon the addition of ND to an anti-parallel G-quadruplex, dramatic changes in the CD spectra were induced (Figure 12), suggesting that ND binds to a human telomeric sequence regardless of the initial secondary structure, and produces a unique structure. Despite the ND binding to the telomeric sequence, a conclusive result regarding the mode of the binding was not obtained because the stoichiometry of the ND binding to the human telomere model sequence d(TTAGGGTTAGGGTTA) (telo15) was not unique. To gain an improved understanding of ND binding to the telomeric sequence, we succeeded in developing a new molecule naphthyridine tetramer (NT) (Figure 5) that binds to telo15 with an exclusive 1:1 binding stoichiometry.[35]

The binding stoichiometry of NT to the telomeric sequence was evaluated by ESI-TOF MS analyses (Figure 13).[49] ESI-TOF MS of telo15 and NT showed a distinct ion at m/z of 1407.71, which corresponded to 4− ion of a 1:1 complex [telo15 + NT]$^{4-}$. Neither ions corresponding to complexes with different binding stoichiometry nor ions of NT bound complexes to telo15 dimer could be detected, even at a higher NT concentration (60 μM). To gain an insight into the mode of NT binding to telo15, ESI-MS of a mutant sequence d(TTAZZZTTAZZZTTA) (telo15Z), where all of the guanine

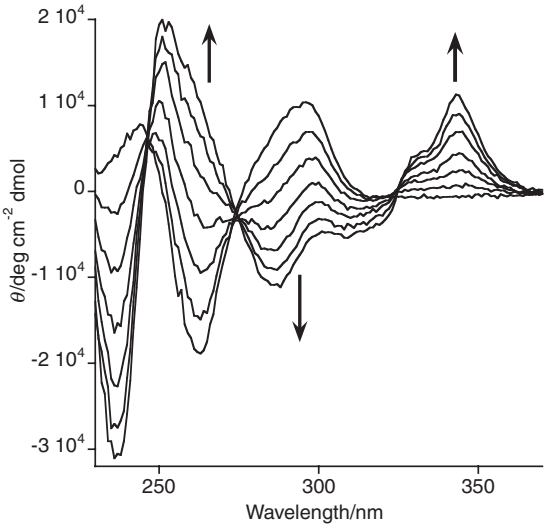


Figure 12. CD spectral change of anti-parallel G-quadruplex of telo22 upon addition of ND. CD measurements were carried out with 5 μM DNA in 100 mM NaCl and 10 mM Na phosphate buffer (pH 7.0) at 7 °C. Concentration of ND was 0, 5, 10, 15, 20, 25, and 30 μM.

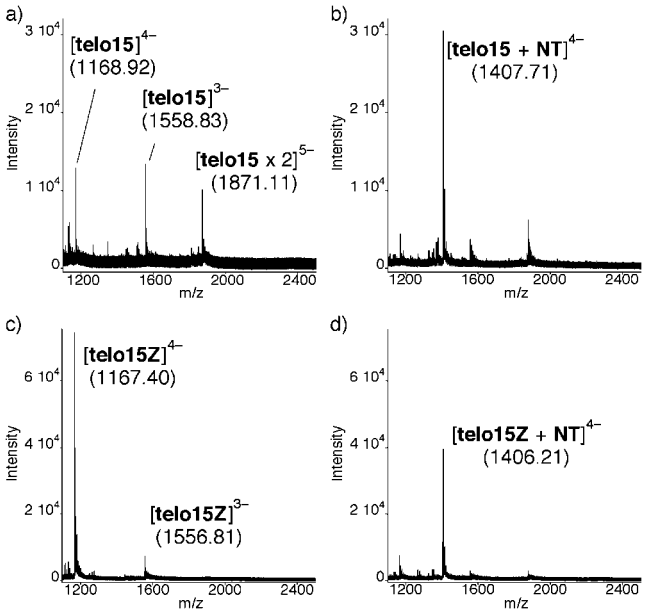


Figure 13. The ESI-TOF MS of telo15 and telo15Z with NT. The spectra were obtained in 50% aqueous methanol and 100 mM NH_4OAc. a) telo15, b) telo15 and NT, c) telo15Z, and d) telo15Z and NT.

base (G) in telo15 were substituted by 7-deazaguanine (Z), and cannot form G-quadruplex (Figure 14). ESI-TOF MS showed a 1:1 complex of telo15Z and NT as of telo15 and NT with the ion at m/z of 1406.21. On the basis of these observations, NT bound to the Watson–Crick face of G and Z. Isothermal titration calorimetry[25] revealed that the binding constant of NT to telo15 was $5.71 \times 10^6\ M^{-1}$ with the titration curve being well fit to a 1:1 binding isotherm.

To know whether the NT-stabilized structures can interfere with the elongation of the telomere sequence, polymerase stop assay[50] was performed on a template containing two- to four-

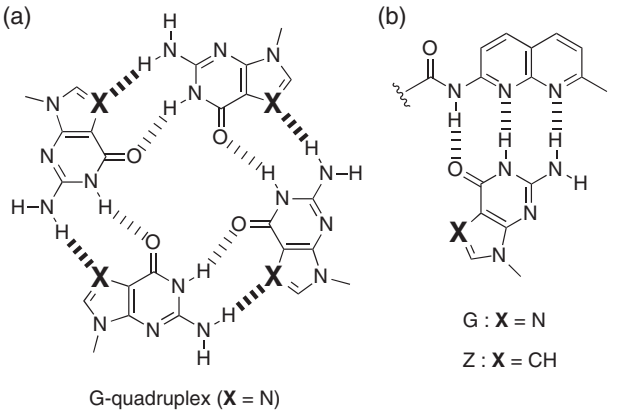


Figure 14. (a) Structure of G-quadruplex (X = N). 7-Deazaguanine (X = CH) cannot form quadruplex due to a lack of hydrogen bonding to the 2-amino group (shown with a bold dash). (b) Hydrogen bonding of *N*-acyl-2-amino-1,8-naphthyridine in NT to guanine (G) and 7-deazaguanine (Z).

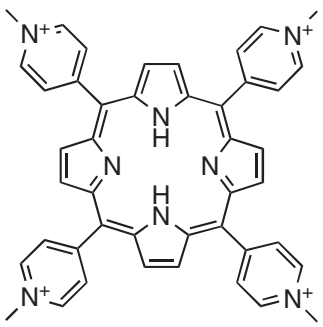


Figure 15. Structure of TMPyP4.

repeat units of d(TTAGGG) sequence as the model telomeric overhang in the presence of NT and a reference ligand TMPyP4[51] (Figure 15).[36] In the absence of ligands, the 20-mer primer that hybridized to the 3′ end of the templates was fully elongated by *Taq* DNA polymerase regardless of the length of telomeric repeats. For the template containing two repeats, NT interfered with the *Taq* polymerase reactions very weakly to produce faint paused bands at the first GGG site from the 3′ end of the repeat (Figure 16).

For the three-repeat template, distinct paused bands were observed predominantly at the first GGG site with NT. The G-quadruplex binding ligand, TMPyP4 did not produce any paused bands on two- and three-repeat templates, but totally suppressed the polymerase reaction at 3 μM. In marked contrast, both NT and TMPyP4 interfered with the *Taq* polymerase reaction on the four-repeat template at the first GGG site. The sequence requirements to form a stable ligand-bound structure and to interfere with the polymerase reaction were quite different for NT and TMPyP4. Four-repeat units were necessary for TMPyP4, showing the formation of TMPyP4-stabilized intrastranded G-quadruplex. The formation of paused bands on the three-repeat template with NT indicated that the NT-bound non-quadruplex structure was stable enough to interfere with the polymerase reaction.

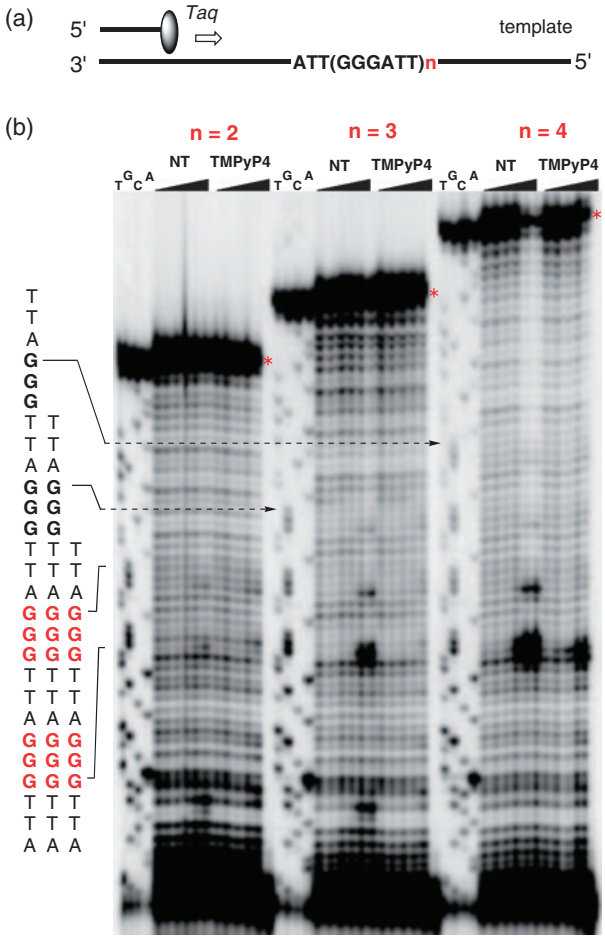


Figure 16. (a) Schematic illustration of a polymerase stop assay of DNA with various numbers of telomeric repeats. (b) Concentration-dependent interruption of *Taq* polymerase-mediated DNA synthesis with NT (0, 0.3, 1, 3, 10, and 30 μM) and TMPyP4 (0, 0.03, 0.1, 0.3, 1, and 3 μM) on DNA templates (0.1 μM) containing (left to right) two, three, and four human telomeric repeats. The lane markers T, G, C, and A indicate the bases on the template strand. The lanes with an asterisk indicated no elongation of the primer. Guanines in the major arrested regions are emphasized in red.

The telomeric overhang in human telomeres is 130–210 nucleotides in length.[52,53] With increasing telomeric repeats, formation of many NT-stabilized hairpin structures in terms of the number of GGG units involved in the hairpin is conceivable. The effects of the repeat length on the NT binding and the interference of the polymerase reaction thereby were investigated with the polymerase stop assay on a template containing up to eight telomeric repeats (Figure 17). With the five-repeat template, the paused bands were observed at the first and the second GGG sites from the 3′ end with an almost equal intensity. In contrast, predominant paused bands at the first GGG site were observed for the six-repeat template. Minor paused bands were detected at the second and third GGG, but not at other GGG sites.

With the seven-repeat template, two major paused bands were detected at the first and second GGG sites. The bands at

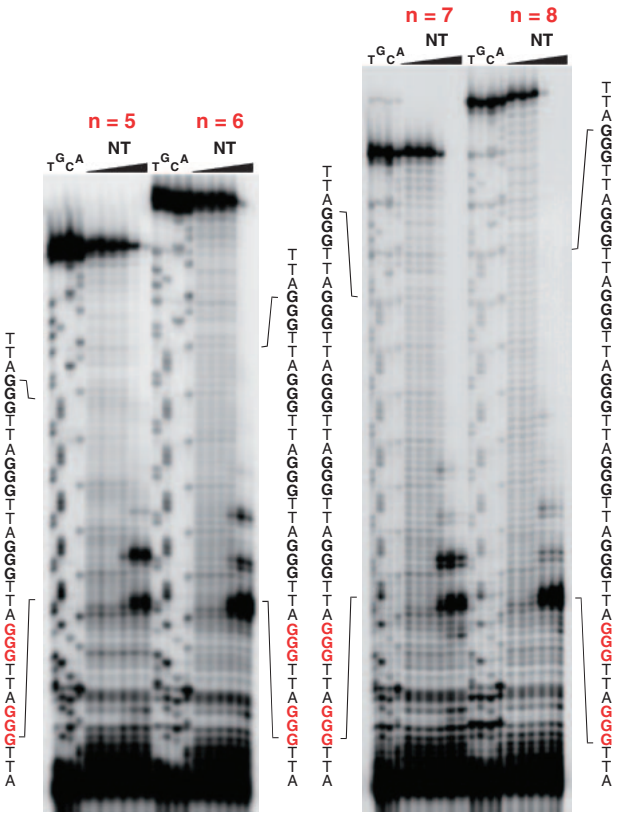


Figure 17. Concentration-dependent inhibition of *Taq* polymerase-mediated DNA synthesis with NT (0, 0.3, 1, 3, 10, and 30 μM) on DNA templates (0.1 μM) containing (left to right) five, six, seven, and eight human telomeric repeats. Guanines in the major arrested sites are emphasized in red.

the second GGG sites, however, gradually faded away with increasing NT concentration with concomitant increase of the band intensity at the first GGG site. With the eight-repeat template, the polymerase reaction was exclusively interrupted at the first GGG site. A threshold concentration for the effective interference became critical on eight-repeat template.

These experiments on five- to eight-repeat templates showed that NT effectively interfered with the polymerase reaction. The site of interruption was repeat dependent, but mostly at the first GGG sites on the template. The interruption of the polymerase reactions becomes sensitive to the NT concentration with increasing repeat numbers to exhibit a clear threshold concentration. These observations regarding the site of and the efficiency for the interference of the polymerase reactions could be well rationalized by the formation of the most stable NT-bound hairpin structure (Figure 18).

4.2 Trinucleotide Repeat Sequences. The expansion of trinucleotide repeats, particularly d(CXG)$_n$, which is common within the human genome is associated with a number of disorders, including Huntington's disease, spinobulbar muscular atrophy, and spinocerebellar ataxia.[54,55] Repeat expansion of d(CAG)$_n$, d(CTG)$_n$, and d(CGG)$_n$ trinucleotides is suggested to relate to the increased stability of alternative DNA hairpin structures consisting of CXG–CXG triads with X–X mismatches.[56,57] Small molecule ligands that selectively bind to d(CXG)$_n$ repeats would be an important probe for determining repeat length and tools for investigating the in vivo repeat extension mechanism.

4.2.1 CAG Trinucleotide Repeats: The number of (CAG)$_n$ repeats in normal IT15 genes ranges from 6 to 39, but is repeated up to 121 times in patients with Huntington's disease.[57] The actual mechanism of repeat expansion remains unclear, but it is likely to involve strand slippage during DNA synthesis mediated by formation of an alternative DNA hairpin structure. The hairpin form of d(CAG)$_n$ repeats involves the

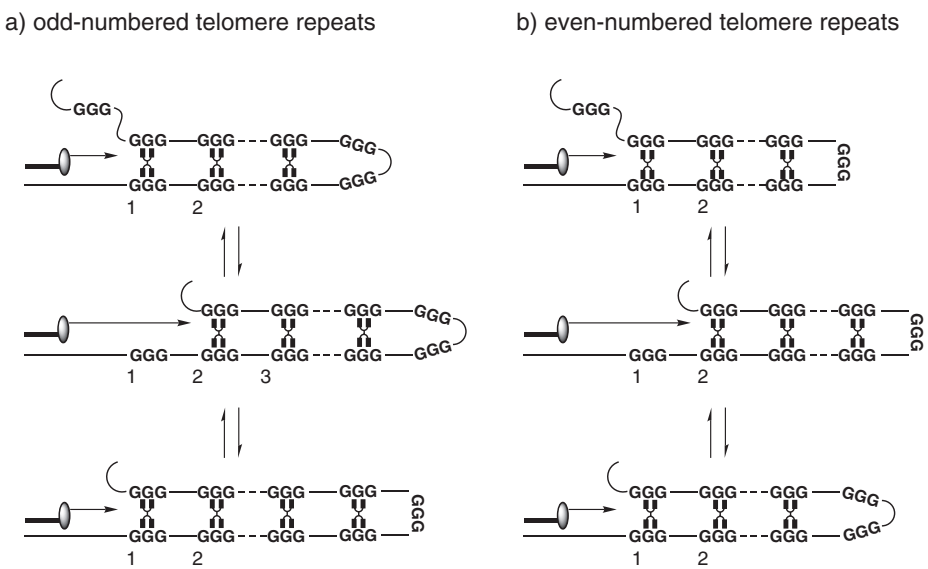


Figure 18. Illustrations for rationalizing the interruption of polymerase by NT-stabilized hairpin structures: a) for odd-numbered telomere repeats and b) for even-numbered repeats. The small number below the GGG units shows the position of the GGG from the 3′-end of the repeat. Experimental results suggested that the hairpin structure in the bottom became most stable with increasing NT-concentration.

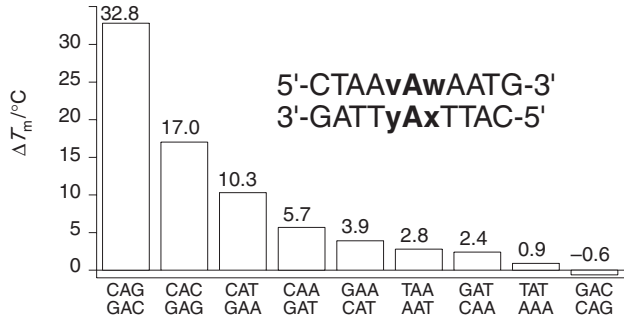


Figure 19. Sequence dependent ΔT_m of NA binding to the A–A mismatch.

intramolecular pairing of CAG–CAG triads with central A–A mismatches being flanked by two G–C base pairs. Thus, ligands binding to the CAG–CAG triad also are expected to bind to the hairpin form of the d(CAG)$_n$ repeat. We discovered that a naphthyridine–azaquinolone NA,[19] which was developed as the ligand for the G–A mismatch recognition strongly bound to the CAG–CAG triad. The thermal stability of an 11-mer duplex 5′-d(CTA ACA GAA TG)-3′/5′-d(CAT TCA GTT AG)-3′ (CAG–CAG), which contained a CAG–CAG triad, was enhanced by 32.8 °C in the presence of NA.[20] NA binding was highly sensitive to the identity of the base pairs flanking the central A–A mismatch (Figure 19). The A–A mismatch in the CAG/CAG triad was stabilized but that in the GAC/GAC triad was not at all, suggesting the importance of the presence of the 5′-side C and/or the 3′-side G to the A–A mismatch. Qualitative analysis of the ΔT_m values suggested that NA binding involved interactions not only with the mismatched A–A pair but also with the 3′-G residue. ESI-TOF mass spectrometry revealed that the stoichiometry of NA binding to the CAG–CAG triad was 2:1. Complexes with other binding stoichiometry were not observed, even under lower or higher NA concentrations. The association constant (K_a) of each NA molecule for the CAG–CAG triad was determined by ITC as $1.8 \times 10^6\,M^{-1}$. The titration curve of the heat produced versus the DNA–ligand ratio supported a binding model involving a single set of identical sites, suggesting that two equivalent molecules of NA bind to and stabilize the CAG–CAG triad.

The solution structures of the NA–CAG–CAG complex determined by NMR reveal how two NA molecules bind to one A–A mismatch and the 3′-G in CAG–CAG. Free CAG–CAG has a canonical B-type DNA conformation and the mismatched adenosine bases are stacked in the helix. In the NA–CAG–CAG complex, two mismatched adenosine bases form intermolecular hydrogen bonds with the 8-azaquinolone units of two NA molecules (Figure 20). NA binds to an adenine base on one strand and a guanine base on the opposite strand, showing that the concept of our molecular design of base recognition through hydrogen bonding was proven feasible. Naphthyridine–guanine and 8-azaquinolone–adenine pairs are well stacked in the right-handed DNA helix, showing structural mimicry of Watson–Crick base pairing. The most extraordinary structural feature of the NA–CAG–CAG complex is the invasion of the G–C base pair by naphthyridine moieties. As a consequence, the two widowed cytidine nucleotides were extruded from the π-stack and produced two C-bulge loops.

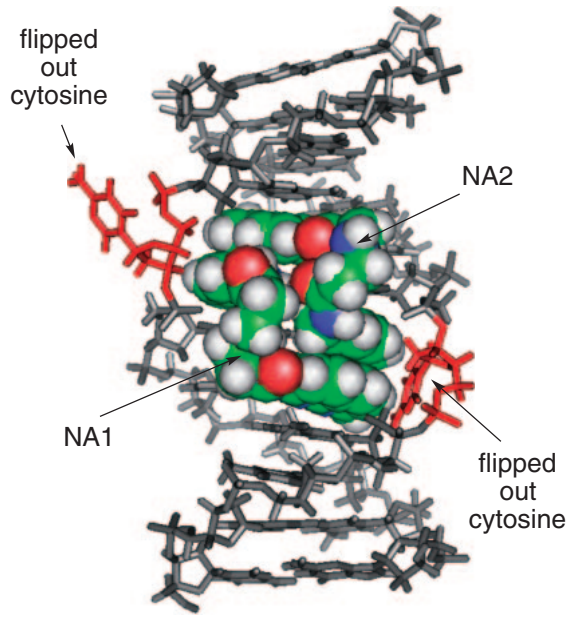


Figure 20. The structure of 2NA–CAG/CAG complex solved by NMR.

The NA–CAG–CAG structure determined by NMR is remarkable because invasion of a small molecule naphthyridine chromophore in NA forced the cytosine to flip out of the helix. To the best of our knowledge, this is the first observation of induction of nucleotide flipping by a small molecular ligand by non-covalent binding. The CAG/CAG triad is the key structural element in the putative hairpin structure involved in the CAG repeats expansion. The NA binding to the CAG repeat sequence was in fact confirmed by CSI-TOF MS as well as the surface plasmon resonance assay with the sensor holding the NA ligand on the surface.

4.2.2 CGG Trinucleotide Repeats: The expansion of the d(CGG)$_n$ trinucleotide repeat in FMR1 gene causes the neurological disorder Fragile X syndrome.[54,55] The molecular basis for the d(CGG)$_n$ expansion involves the formation of a metastable hairpin structure[56,57] consisting of continued 5′-d(CGG)-3′/5′-d(CGG)-3′ triads, where a G–G mismatch is flanked by two G–C base pairs. Since we have reported a series of ligands binding to a G–G mismatch, the remarkable NA-bound structure to the CAG/CAG triad[20] prompted us to investigate the mode of the ligand binding to the CGG/CGG triad and hence, the possibility of the cytosine flipping out in the ligand-bound complex (Figure 21). We found that naphthyridine carbamate dimer (NCD)[26] binds to a single CGG/CGG triad with exclusively 2:1 NCD-triad stoichiometry.[27] The NCD binding to the CGG/CGG triad induced disruption of the guanine–cytosine base pairing in the triad and made the cytosine susceptible to the subsequent chemical cleavage reaction initiated by hydroxylamine addition.

The NCD binding to the CGG/CGG triad in the 13-mer duplex increased the T_m by 23.1 °C, whereas, ΔT_m was only 6.7 °C for the GGC/GGC triad (Table 6). A survey of the effect of the flanking sequence suggested that the strong and selective NC binding to the CGG/CGG triad is due to the interaction of NCD not only to the G–G mismatch but also to the 3′ side

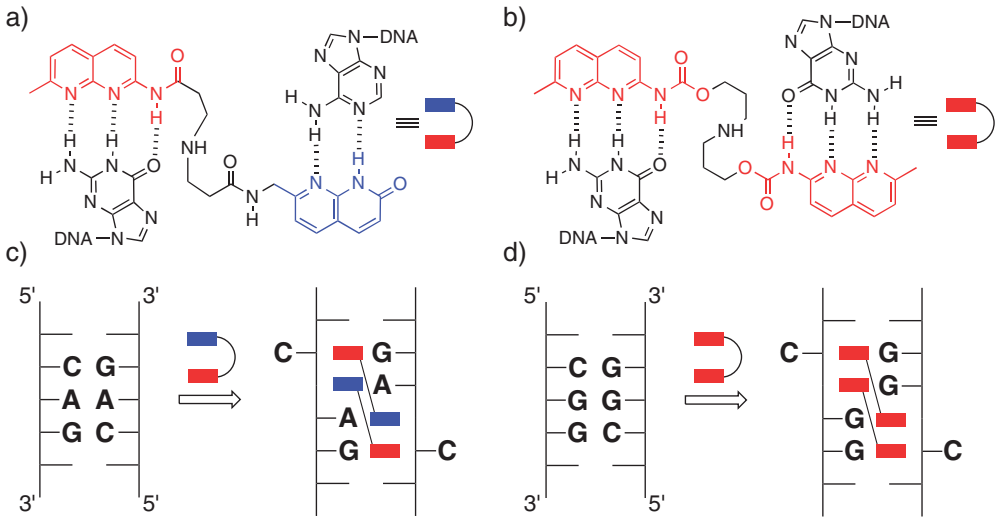


Figure 21. Hydrogen-bonding schemes between a) NA and G–A mismatch and b) NCD and G–G mismatch. Illustrations of c) the NMR-confirmed NA–CAG/CAG triad complex and d) a proposed NCD binding to CGG/CGG triad. Red squares represent 2-amino-1,8-naphythyridine, whereas blue squares indicate 8-azaquinolone.

Table 6. ΔT_m Values for the 13-mer Duplexes Containing a G–G Mismatch in a Different Flanking Sequence[a)]

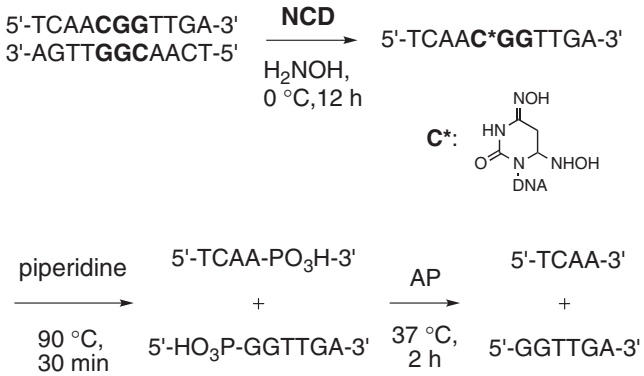
5′-GCTAA**XGZ**AATGA-3′
3′-CGATT**YGW**TTACT-5′

XGZ	YGW	$T_{m(-)}$	ΔT_m
CGG	GGC	34.1 (0.9)	23.1 (0.4)
CGC	GGG	38.6 (0.1)	10.7 (0.8)
GGC	CGG	40.4 (0.3)	6.7 (1.4)
CGA	GGT	31.8 (0.2)	12.6 (0.3)
CGT	GGA	33.6 (0.2)	10.4 (0.5)
GGA	CGT	34.2 (0.3)	9.2 (0.4)
AGT	TGA	28.7 (0.4)	13.2 (1.0)
AGG	TGC	31.8 (0.6)	8.1 (0.8)
GGT	CGA	33.7 (0.4)	5.6 (0.3)
TGA	AGT	17.8 (0.4)	18.1 (1.0)

a) The UV-melting curve was measured for a duplex (4.5 μM) in a 10 mM sodium cacodylate buffer (pH 7.0) containing 100 mM NaCl and NCD (100 μM). Temperature was increased at a rate of 1 °C min^{-1}. All measurements were taken three times, and standard deviations are shown in parentheses.

guanine. CSI-TOF MS clearly demonstrated that the binding of NCD to the duplex containing the CGG/CGG triad proceeded in an exclusive stoichiometry of 2:1. The crossover point of the Job's plot obtained from the UV titration lied at 66%, confirming that the binding stoichiometry of NCD to the CGG/CGG triad would be 2:1 as determined by CSI-TOF MS.

The high sequence preference and the 2:1 binding stoichiometry for the NCD binding are the characteristic features that we observed for the NA binding to the CAG/CAG triad.[20] We looked at the reaction of NCD-bound CGG/CGG triad with hydroxylamine if the cytosine in the CGG/CGG triad was free from the base pairing upon NCD-binding. Cytosines in the single-stranded regions and in the mismatched base pairs efficiently reacted with two molecules of hydroxylamine at C4 and C6 positions, whereas those in the G–C base pair in a duplex were not susceptible to the addition of hydroxylamine.

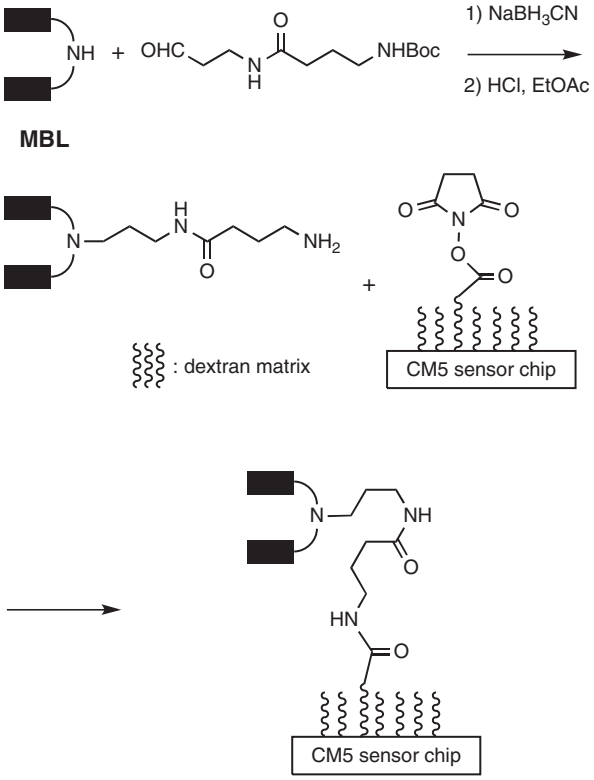


Scheme 1. Hydroxylamine induced cleavage at cytosine in the NCD-bound CGG/CGG triad.

The cytosine modified with hydroxylamine undergoes degradation by heating with piperidine eventually leading to a strand cleavage.[58] The HPLC and MALDI-TOF analyses of the reaction of 11-mer self-complementary 5′-d(TCAA **CGG** TTGA)-3′ with hydroxylamine clarified that the cytosines in the CGG/CGG triad in the duplex became susceptible to the hydroxylamine upon binding with NCD (Scheme 1), suggesting that the NCD binding to the CGG/CGG triad made the cytosine free from the base pairing to the guanine in the opposite strand. The NCD binding to the CGG repeat sequence was also confirmed by CSI-TOF MS. The CD spectra of $d(CGG)_{10}$ showed a large conformational change upon NCD binding, strongly suggesting the induction of hairpin secondary structure in the repeat sequence.

5. Potential Application of Mismatch Binding Ligands

The MBL we have developed is a key molecular element in developing innovative tools and devices for the detection of single nucleotide polymorphisms (SNPs)[13,19,21,33,59] and for the determination of the length of repeat sequences.[20,27] SNP is the single nucleotide difference of two DNA sequence and the origin of individual difference in phenotype.[60] Facile



Scheme 2. Preparation of MBL-immobilized surface.

SNP detection is necessary technology to realize personalized medicine. When two DNAs were denatured and re-annealed, the site of SNP if existed would appear as a mismatched base pair. Detection of the mismatched base pairs implies the detection of SNPs. This is the basis of heteroduplex analysis that has been conventionally used for the detection of genetic mutations.[61] We have applied the MBL for the detection of SNPs in heteroduplex analysis. One approach we have taken is a surface plasmon resonance assay using the sensor holding the MBL on its surface. The other approach is affinity separation with a MBL–Sepharose column.

5.1 MBL–SPR Sensor for Mismatch Detection. We planned to use surface plasmon resonance (SPR) assay for the detection of mismatched DNA by using a sensor chip holding the MBL on the surface. As we discussed the molecular design of MBL, MBL has a secondary amino group in the linker. The secondary nitrogen is the site of a tether to another short linker, which serves as the site for covalent immobilization of MBL on the surface of the solid support. For the connection of a short linker, an amide bond formation was initially used. However, an MBL with a tertiary amide in the linker was not as effective as the original molecule toward mismatch binding.[14] This was most likely due to the rigid conformation of the linker moiety. Therefore, the short linker was attached by reductive amination to preserve the basic nitrogen and the conformational flexibility (Scheme 2). ND was immobilized through a short linker to the carboxyl group on the SPR sensor chip. The sensorgrams were obtained by analyzing 27-mer duplex DNA containing a G–G, G–A, or G–T mismatch in the flanking sequence of 5′CGG3′/5′CXG3′ (X = G, A, and T), using the ND-tethered SPR sensor (Figure 22a).[13] A marked SPR response was obtained only for

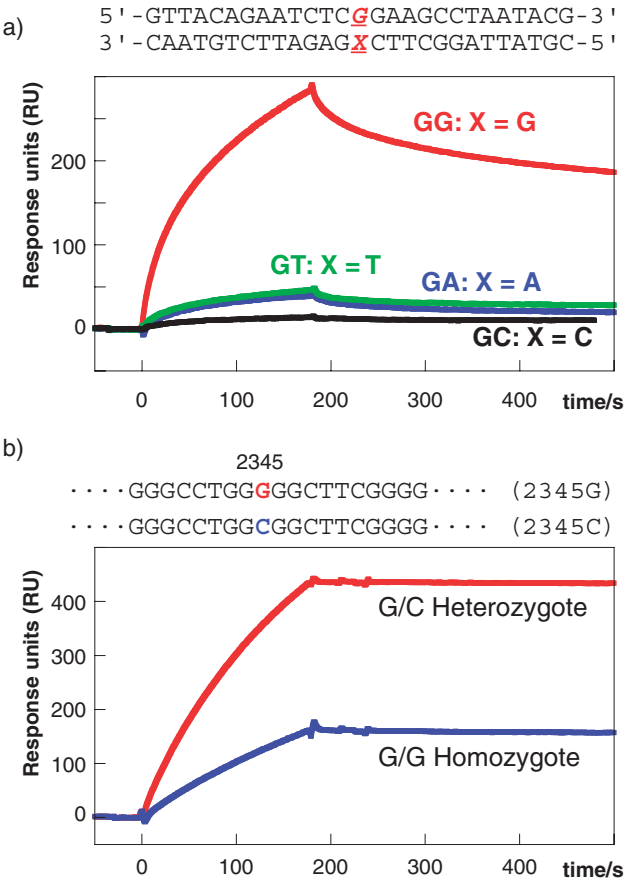


Figure 22. SPR analysis by ND–SPR sensor. a) Analysis of 33-mer duplex having G–X (X = G, A, T, and C) base pair. b) Analysis of 652-nucleotide sequence obtained by PCR amplification of the HSP70-2 gene between the nucleotide position 1986 and 2637. A SNP between G/G homozygote and G/C heterozygote is present at a nucleotide position of 2345 of the gene.

the duplex containing the G–G mismatch, whereas duplexes containing G–A and G–T mismatches, as well as a fully matched duplex, produced only a weak response.

To verify detection of G–G mismatches in real biological samples, the detection of a predetermined SNP between G/G homozygote and G/C heterozygote at a nucleotide position of 2345 in the HSP70-2 gene (GenBank accession number M59830) was demonstrated. PCR amplification of two genes of G/C heterozygote, one containing G and the other containing C at a nucleotide position of 2345, was carried out between nucleotide positions of 1986 and 2537. According to the established method for heteroduplex analysis, PCR products of 652 base pairs were heat-denatured and re-annealed, and analyzed by ND-immobilized SPR sensor. PCR products derived from G/C heterozygote showed much stronger SPR response than that from G/G homozygote (Figure 22b).

For the C–C mismatch detection, the SPR chips holding amND on the sensor surface were prepared. SPR analyses of mismatch-containing 27-mer 5′-d(GTT ACA GAA TCT GXC AAG CCT AAT ACG)-3′/3′-d(CAA TGT TTC AGA CYG TTC GGA TTA TGC)-5′ were performed with amND-immobilized sensor surface at pH 7.0.[21,59] A marked SPR response

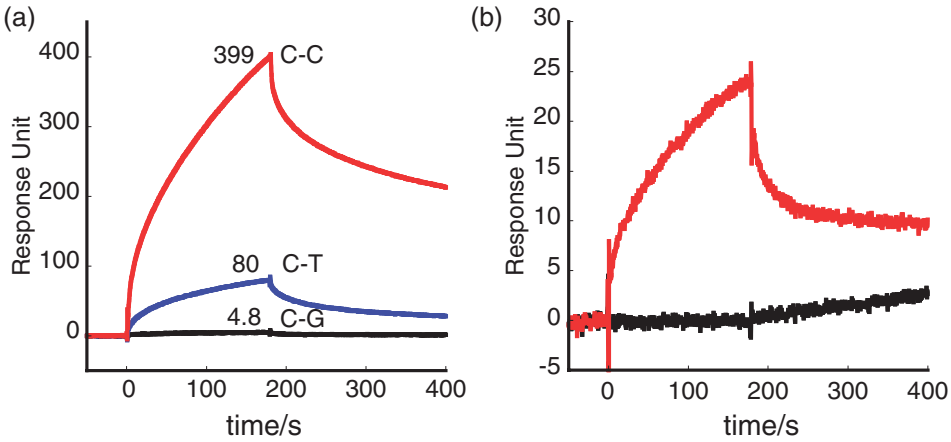


Figure 23. (a) SPR assay of 27-mer duplex (**GXC**/**CYG**) (0.2 μM) containing C–C (**GCC**/**CCG**) (red), C–T (**GCC**/**CTG**) (blue), and C–G (**GCC**/**CGG**) (black) with the amND-immobilized sensor surface. Binding was measured for 180 s and dissociation for 220 s in a phosphate buffer (200 mM, pH 7.0) containing NaCl (150 mM). (b) SPR assay of C–C mismatch (red), C–G match (black) at 10 nM.

was obtained for the duplex containing the C–C mismatch at the DNA concentration of 200 nM (Figure 23a). The response at 180 s after an injection of DNA was 399 RU, about 83 times stronger than that obtained for the fully matched duplex (4.8 RU). The duplex containing a C–T mismatch produced an intermediate response (80 RU). With the amND sensor, the C–C mismatch was detected at 10 nM (Figure 23b). Further lowering the concentration of the C–C mismatch resulted in a loss of signals.

5.2 MBL–Sepharose for Affinity Separation. Among the many technologies that have been developed for SNP typing,[60] methods that separate DNA on a stationary phase, e.g., single-stranded conformational polymorphisms have been used in mutation detection.[62] In these methods using a stationary phase, the chemical basis for the separation was ascribed to the decreased stability of the mismatched duplex compared to the fully matched duplex. We have taken an alternative approach to improve separation of mismatched duplexes by exploiting the binding of MBL. Three MBL–Sepharose columns were prepared by immobilizing the MBL to the NHS-activated Sepharose column.[63] The hairpin duplexes 5′-(CTAACxGAA-TGTTTTCATTCyGTTAG)-3′ containing a mismatched base pair (x–y) in the middle of the 11-mer stem region were analyzed by the three MBL–Sepharose columns. The DNAs were eluted with a gradient of 50 mM NaOH. The elution profiles of mismatched duplexes and a fully matched duplex on the three MBL–Sepharose columns are shown in Figure 24.

The G–G and G–A mismatched DNA were strongly retained on the ND–Sepharose column, and eluted from the column by flowing 50 mM NaOH. The retention times (t_R) for G–G and G–A mismatched duplexes were 9.12 and 6.03 min, respectively (Figure 24a). Under the same conditions, the rest of the six mismatched and a fully matched duplex were eluted within 3 min from the column without elution of 50 mM NaOH due to their weak binding to the surface (Figure 24b). However, the t_R for C–C and G–T mismatches and the peak width at half height of these mismatches suggested that C–C and G–T mismatches were separated by a weak interaction with the ND–Sepharose surface. The elution profiles of the mismatched duplexes obtained with the NA–Sepharose column were quite different from those with ND–Sepharose column. Four mismatched duplexes (A–A, G–A, G–G, and A–C) were retained on the NA–Sepharose column and eluted off upon increasing the fraction of 50 mM NaOH (Figure 24c). The other four mismatched duplexes (G–T, C–C, T–C, and T–T) and the matched duplex were eluted off without NaOH. With the amND–Sepharose column, both the fully matched and mismatched duplexes were retained on the column and were eluted by increasing the fraction of 50 mM NaOH. The C–C mismatch was separated from other mismatches, although the difference in t_R was small (Figure 24d).

Three MBL–Sepharose columns showed characteristic elution profiles for the mismatched duplexes. The elution order of the mismatch was in good agreement with the affinity of each MBL for the mismatched duplexes. These studies described here demonstrated that the MBL-immobilized surface could be useful for the analysis and separation of mismatched DNAs and eventually for SNP typing.

6. Perspective of Mismatch Binding Ligands

Mismatch binding ligands discussed here are a particularly important class of compounds that can be effectively used in detection and analysis of mismatched base pairs in DNA and applicable in SNP detection and typing.[64] We have so far discovered MBL targeting G–G, G–A, and C–C mismatches in duplex DNA. The effect of flanking base pairs on the MBL binding to the mismatched base pairs is quite significant. Thus, the molecular design of the next generation of MBL will be necessary to take into account the information of flanking base pairs of mismatches.

Besides application in genomic science, MBL shows promising possibilities in applying material science using DNA as a unit of nano-scaled structure. The 2:1 binding motifs discussed in CAG and CGG repeat recognition suggested that the flipped out nucleotide bases could be replaced by other nucleotide bases. This inspired us to propose the concept “molecular glue for DNA” that controls the DNA hybridization

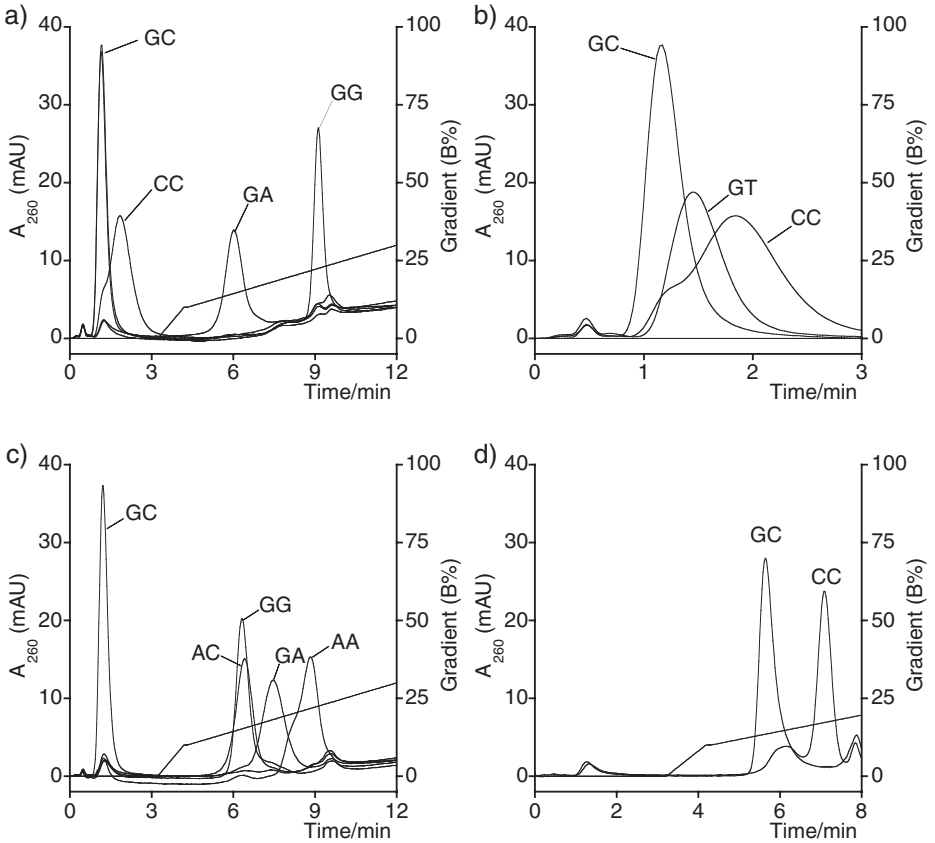


Figure 24. The elution profile of hairpin duplexes containing a mismatched base pair using (a and b) ND–, (c) NA–, and (d) amND–Sepharose columns. The sample DNAs (5 μM) were analyzed by flowing a solvent mixture of elution buffer A (100 mM NaCl, 10 mM Na phosphate, pH 7.0) and elution buffer B (100 mM NaCl, 50 mM NaOH) at a flow rate 1.0 mL min^{-1}. Elution buffer B composition (%) is also plotted. UV absorbance at 260 nm of the eluent was monitored.

by ligand binding.[28,65] We have succeeded in turning ON and OFF DNA hybridization by using photo switching molecular glue for DNA by external light irradiation.[66]

Most importantly, these studies described here clearly showed the possibility to rationally design the molecule that can precisely recognize the base sequences and structures on the basis of the knowledge on organic chemistry and nucleic acid chemistry. Such molecules have promises to open the new view of genomic science, molecular biology, and biotechnology in the future.

The author sincerely thanks collaborators who drive the studies forward and make this research quite fruitful. Their names were shown in the references. The author also thanks Prof. Isao Saito for his encouragement and support. Financial support from the Ministry of Education, Culture, Sports, Science and Technology, the Ministry of Health, Labor and Welfare, and the Japan Science and Technology Agency are deeply acknowledged.

References

1 P. B. Dervan, *Bioorg. Med. Chem.* **2001**, *9*, 2215.
2 P. B. Dervan, B. S. Edelson, *Curr. Opin. Struct. Biol.* **2003**, *13*, 284.
3 T. Bando, H. Sugiyama, *Acc. Chem. Res.* **2006**, *39*, 935.
4 T. Lindahl, *Nature* **1993**, *362*, 709.
5 I. Saito, K. Nakatani, *Bull. Chem. Soc. Jpn.* **1996**, *69*, 3007.
6 R. P. Sinha, D.-P. Häder, *Photochem. Photobiol. Sci.* **2002**, *1*, 225.
7 H. E. Krokan, R. Standal, G. Slupphaug, *Biochem. J.* **1997**, *325*, 1.
8 T. A. Kunkel, K. Bebenek, *Annu. Rev. Biochem.* **2000**, *69*, 497.
9 A. Sancar, *Annu. Rev. Biochem.* **1996**, *65*, 43.
10 R. J. Roberts, X. Cheng, *Annu. Rev. Biochem.* **1998**, *67*, 181.
11 L. S. Kappen, I. H. Goldberg, *Biochemistry* **1992**, *31*, 9081.
12 H. Junicke, J. R. Hart, J. Kisko, O. Glebov, I. R. Kirsch, J. K. Barton, *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 3737.
13 K. Nakatani, S. Sando, I. Saito, *Nat. Biotechnol.* **2001**, *19*, 51.
14 K. Nakatani, S. Sando, H. Kumasawa, J. Kikuchi, I. Saito, *J. Am. Chem. Soc.* **2001**, *123*, 12650.
15 K. Nakatani, S. Sando, I. Saito, *J. Am. Chem. Soc.* **2000**, *122*, 2172.
16 P. S. Corbin, S. C. Zimmerman, *J. Am. Chem. Soc.* **1998**, *120*, 9710.
17 T. J. Murray, S. C. Zimmerman, *J. Am. Chem. Soc.* **1992**, *114*, 4010.
18 T. R. Kelly, C. Zhao, G. J. Bridger, *J. Am. Chem. Soc.* **1989**, *111*, 3744.
19 S. Hagihara, H. Kumasawa, Y. Goto, G. Hayashi, A.

Kobori, I. Saito, K. Nakatani, *Nucleic Acids Res.* **2004**, *32*, 278.
20 K. Nakatani, S. Hagihara, Y. Goto, A. Kobori, M. Hagihara, G. Hayashi, M. Kyo, M. Nomura, M. Mishima, C. Kojima, *Nat. Chem. Biol.* **2005**, *1*, 39.
21 A. Kobori, S. Horie, H. Suda, I. Saito, K. Nakatani, *J. Am. Chem. Soc.* **2004**, *126*, 557.
22 K. Nakatani, H. He, S. Uno, T. Yamamoto, C. Dohno, *Current Protocols in Nucleic Acid Chemistry,* **2008**, Unit 8.6, p. 1.
23 K. Nakatani, S. Sando, I. Saito, *Bioorg. Med. Chem.* **2001**, *9*, 2381.
24 M. Brenowitz, D. F. Senear, M. A. Shea, G. K. Ackers, *Methods Enzymol.* **1986**, *130*, 132.
25 I. Haq, J. O. Trent, B. Z. Chowdhry, T. C. Jenkins, *J. Am. Chem. Soc.* **1999**, *121*, 1768.
26 T. Peng, T. Murase, Y. Goto, A. Kobori, K. Nakatani, *Bioorg. Med. Chem. Lett.* **2005**, *15*, 259.
27 T. Peng, K. Nakatani, *Angew. Chem., Int. Ed.* **2005**, *44*, 7280.
28 T. Peng, C. Dohno, K. Nakatani, *Angew. Chem., Int. Ed.* **2006**, *45*, 5623.
29 F. H. Beijer, H. Kooijman, A. L. Spek, R. P. Sijbesma, E. W. Meijer, *Angew. Chem., Int. Ed.* **1998**, *37*, 75.
30 J. Pranata, S. G. Wierschke, W. L. Jorgensen, *J. Am. Chem. Soc.* **1991**, *113*, 2810.
31 J. Sartorius, H.-J. Schneider, *Chem.—Eur. J.* **1996**, *2*, 1446.
32 H. Suda, A. Kobori, J. Zhang, G. Hayashi, K. Nakatani, *Bioorg. Med. Chem.* **2005**, *13*, 4507.
33 F. Takei, H. Suda, M. Hagihara, J. Zhang, A. Kobori, K. Nakatani, *Chem.—Eur. J.* **2007**, *13*, 4452.
34 K. Nakatani, S. Hagihara, S. Sando, S. Sakamoto, K. Yamaguchi, C. Maesawa, I. Saito, *J. Am. Chem. Soc.* **2003**, *125*, 662.
35 Y. Goto, S. Hagihara, M. Hagihara, K. Nakatani, *ChemBioChem* **2007**, *8*, 723.
36 M. Hagihara, Y. Goto, K. Nakatani, *ChemBioChem* **2008**, *9*, 510.
37 E. A. Smith, M. Kyo, H. Kumasawa, K. Nakatani, I. Saito, R. M. Corn, *J. Am. Chem. Soc.* **2002**, *124*, 6810.
38 C. W. Greider, *Annu. Rev. Biochem.* **1996**, *65*, 337.
39 E. H. Blackburn, *Nature* **1991**, *350*, 569.
40 C. B. Harley, A. B. Futcher, C. W. Greider, *Nature* **1990**, *345*, 458.
41 N. D. Hastie, M. Dempster, M. G. Dunlop, A. M. Thompson, D. K. Green, R. C. Allshire, *Nature* **1990**, *346*, 866.
42 S. Neidle, G. Parkinson, *Nat. Rev. Drug Discovery* **2002**, *1*, 383.
43 L. H. Hurley, *Nat. Rev. Cancer* **2002**, *2*, 188.
44 Y. Xu, Y. Noguchi, H. Sugiyama, *Bioorg. Med. Chem.* **2006**, *14*, 5584.
45 A. Ambrus, D. Chen, J. Dai, T. Bialis, R. A. Jones, D. Yang, *Nucleic Acids Res.* **2006**, *34*, 2723.
46 K. N. Luu, A. T. Phan, V. Kuryavyi, L. Lacroix, D. J. Patel, *J. Am. Chem. Soc.* **2006**, *128*, 9963.
47 Y. Wang, D. J. Patel, *Structure* **1993**, *1*, 263.
48 J. R. Williamson, *Curr. Opin. Struct. Biol.* **1993**, *3*, 357.
49 J. L. Beck, M. L. Colgrave, S. F. Ralph, M. M. Sheil, *Mass Spectrom. Rev.* **2001**, *20*, 61.
50 H. Han, L. H. Hurley, M. Salazar, *Nucleic Acids Res.* **1999**, *27*, 537.
51 R. T. Wheelhouse, D. Sun, H. Han, F. X. Han, L. H. Hurley, *J. Am. Chem. Soc.* **1998**, *120*, 3261.
52 V. L. Makarov, Y. Hirose, J. P. Langmore, *Cell* **1997**, *88*, 657.
53 W. E. Wright, V. M. Tesmer, K. E. Huffman, S. D. Levene, J. W. Shay, *Genes Dev.* **1997**, *11*, 2801.
54 C. T. Ashley, Jr., S. T. Warren, *Annu. Rev. Genet.* **1995**, *29*, 703.
55 *Genetic Instabilities and Hereditary Neurological Diseases*, ed. by R. D. Wells, S. T. Warren, Academic Press, San Diego, **1998**.
56 A. M. Gacy, G. Goellner, N. Juranić, S. Macura, C. T. McMurray, *Cell* **1995**, *81*, 533.
57 M. Mitas, A. Yu, J. Dill, T. J. Kamp, E. J. Chambers, I. S. Haworth, *Nucleic Acids Res.* **1995**, *23*, 1050.
58 B. H. Johnston, *Methods Enzymol.* **1992**, *212*, 180.
59 A. Kobori, K. Nakatani, *Bioorg. Med. Chem.* **2008**, *16*, 10338.
60 K. Nakatani, *ChemBioChem* **2004**, *5*, 1623.
61 A. J. Nataraj, I. Olivos-Glander, N. Kusukawa, W. E. Highsmith, Jr., *Electrophoresis* **1999**, *20*, 1177.
62 K. Hayashi, D. W. Yandell, *Hum. Mutat.* **1993**, *2*, 338.
63 Y. Goto, H. Suda, A. Kobori, K. Nakatani, *Anal. Bioanal. Chem.* **2007**, *388*, 1165.
64 T. Peng, H. He, M. Hagihara, K. Nakatani, *ChemBioChem* **2008**, *9*, 1893.
65 T. Peng, C. Dohno, K. Nakatani, *ChemBioChem* **2007**, *8*, 483.
66 C. Dohno, S. Uno, K. Nakatani, *J. Am. Chem. Soc.* **2007**, *129*, 11898.

Kazuhiko Nakatani was born in 1959 and received his B.Sc. (1982) and Ph.D. (1988) from Osaka City University. During his Ph.D. studies, he worked in the laboratory of Professor Gilbert Stork of Columbia University for three years. He spent three years as a postdoc at Sagami Chemical Research Center under the guidance of Dr. Shiro Terashima. He started his academic career at Osaka City University as Assistant Professor in 1991. After two years, he joined Professor Isao Saito's group in Kyoto University, where he learned much about nucleic acid chemistries. He was promoted to Associate Professor in 1997, then Professor of Osaka University in 2005. His research interests are mainly in genomic science based on organic chemistry. He is a recipient of The Japan IBM award (2005), Ichimura Academic award (2008), Osaka Science award (2008), and The Chemical Society of Japan Award for Creative Work for 2007.