Review

Possibilities and challenges of small molecule organic compounds for the treatment of repeat diseases

By Kazuhiko NAKATANI[*1,†]

(Edited by Keisuke SUZUKI, M.J.A.)

Abstract: The instability of repeat sequences in the human genome results in the onset of many neurological diseases if the repeats expand above a certain threshold. The transcripts containing long repeats sequester RNA binding proteins. The mechanism of repeat instability involves metastable slip-out hairpin DNA structures. Synthetic organic chemists have focused on the development of small organic molecules targeting repeat DNA and RNA sequences to treat neurological diseases with repeat-binding molecules. Our laboratory has studied a series of small molecules binding to mismatched base pairs and found molecules capable of binding CAG repeat DNA, which causes Huntington's disease upon expansion, CUG repeat RNA, a typical toxic RNA causing myotonic dystrophy type 1, and UGGAA repeat RNA causing spinocerebellar ataxia type 31. These molecules exhibited significant beneficial effects on disease models *in vivo*, suggesting the possibilities for small molecules as drugs for treating these neurological diseases.

Keywords: small molecule, neurological diseases, mismatch, trinucleotide repeat, hairpin, expansion

Introduction

The human genome contains many repeat sequences related to human genetic diseases.[1] Telomere repeats (TTAGGG) at the ends of human chromosomes shorten as the genome replicates,[2]–[4] and when the repeat length falls below a certain threshold, cell death is induced. In cancer cells that have acquired immortality, the enzyme telomerase is activated to elongate the shortened telomere repeat.[5]

Depending on their genomic location, sequence, and the length of repeats, repeat sequences are associated with the development of various neurodegenerative diseases.[6] Repetitions of aberrantly expanded trinucleotide sequences (trinucleotide repeats) are known to be the cause of trinucleotide repeat diseases.[7],[8]

Specifically, the CGG repeat sequence in exon 1 of the *FMR1* gene on the X chromosome is approximately 40 repeats in healthy individuals, but when it exceeds 200 repeats due to aberrant expansion, Fragile X Syndrome develops.[9]–[11] (Table 1) Similarly, the CCG repeat sequence on the *FMR2* gene on the X chromosome is approximately 40 repeats in normal individuals but exceeds 200 repeats in patients with Fragile XE syndrome.[12]–[14] The CGG repeat sequence in exon 1 of the *FMR1* gene and the CCG repeat sequence in the *FMR2* gene are both located in the 5′-untranslated region (5′-UTR) of their respective genes. Huntington's disease (HD) is caused by an expansion of CAG repeats in the coding region of the first exon of the *HTT* gene, which encodes the huntingtin protein on chromosome 4.[15]–[17] The length of the CAG repeat is approximately 35 repeats or less in healthy individuals, but when it exceeds 40 repeats,

[*1] SANKEN, The Institute of Scientific and Industrial Research, Osaka University, Ibaraki, Osaka, Japan.

[†] Correspondence should be addressed: K. Nakatani, SANKEN, The Institute of Scientific and Industrial Research, Osaka University, 8-1 Mihogaoka, Ibaraki, Osaka 567-0074, Japan (e-mail: nakatani@sanken.osaka-u.ac.jp).

Non-standard abbreviation list: 5′-UTR: 5′-untranslated region; ΔT_m: change in melting temperature; CD: circular dichroism; DANP: 2,7-diamino-1,8-naphthyridine; DM1: myotonic dystrophy type 1; EMSA: electrophoretic mobility shift assay; HD: Huntington's disease; MBL: mismatch binding ligand; NA: naphthyridine-azaquinolone; NCD: naphthyridine carbamate dimer; NMR: nuclear magnetic resonance; QCD: quinoline carbamate dimer; R-loop: DNA-RNA hybrids; RBP: RNA-binding protein; S-DNA: slipped strand DNA; SCA31: spinocerebellar ataxia type 31; SPR: surface plasmon resonance; T_m: melting temperature; UDG: uracil DNA glycosylase.

Table 1. Repeat sequences related to neurological disorders

Repeat sequence	Gene (location)	Gene disorder	Wild-type (n =)	Mutant type (n =)
$(CGG)_n$	*FMR1* (Chr. X, exon 1, 5′-UTR)	Fragile X syndrome	5–40	60–200
$(CCG)_n$	*FMR2* (Chr. X, exon 1, 5′-UTR)	Fragile XE syndrome	4–39	>200
$(CAG)_n$	*HTT* (Chr. 4, exon 1, ORF)	Huntington disease	6–39	36–121
$(GAA)_n$	*X25 (frataxin)* (Chr. 9, intron 1)	Friedreich ataxia	6–34	112–1700
$(CTG)_n$	*DMPK* (Chr. 19, exon 15, 3′-UTR)	Myotonic dystrophy type 1	5–37	50–6500
$(TGGAA)_n$	*BEAN, TK2* (Chr. 19, intron)	Spinocerebellar ataxia type 31	0	500–760
$(GGGGCC)_n$	*C9orf72* (Chr. 9, intron)	Amyotrophic lateral sclerosis	2–19	250–1600

the probability of developing HD increases. The CAG repeats are translated into glutamine, so the aberrantly expanded CAG repeats form a long continuous region of glutamine in the huntingtin protein. Aggregation of the glutamine region was thought to be the cause of the disease onset,[18)–20)] HD is also called polyglutamine disease. Myotonic dystrophy type 1 (DM1) is a systemic disease characterized by muscle atrophy, myotonia, testicular atrophy, cataracts, and glucose intolerance.[21)–23)] The gene responsible for DM1 is the *DMPK* gene on chromosome 19, and the direct cause of the disease is the aberrant expansion of CTG repeats in its 3′-UTR. The number of repeats in normal individuals ranges from 5 to 37, and DM1 is diagnosed when the number of repeats exceeds 50.

Disease-causing repeat sequences are not limited to trinucleotide sequences; diseases involving penta and hexanucleotide repeat sequences are also known. Spinocerebellar ataxia type 31 (SCA31)[24)] is caused by the insertion of a TGGAA pentanucleotide repeat extending 2.5 to 3.8 kilobases into an intron shared by the *BEAN* and *TK2* genes on chromosome 16.[25)] In addition, approximately 40% of familial amyotrophic lateral sclerosis is reported to be caused by aberrant expansion of the GGGGCC hexanucleotide repeat sequence in intron 1 of *C9orf72*.[26),27)] Expansion or insertion of these repeats directly affects the amino acid sequence of the protein if the repeats are in the translated region, as in the *HTT* gene. In addition, many repeat diseases are now thought to be

caused by aberrantly long repeat sequences in transcripts, whether translated or not, that acquire the ability to capture RNA-binding proteins. The term "toxic RNA" is used to describe RNA that has acquired toxicity.[28)–35)]

Repeat sequences in the genome can expand or shorten during biological reactions such as replication and transcription, where the double strands dissociate.[36),37)] Aberrantly expanded repeat sequences are known to form metastable slipped strand DNA (S-DNA) with a hairpin structure in each repeat region on both strands when the double strand is dissociated, and the stability of the S-DNA structure is thought to be related to the ease of repeat expansion and shortening.[38)–40)] (Fig. 1a) Sequestration of RNA-binding proteins by toxic RNA also involved the hairpin structure of the expanded RNA repeat. For example, in the aberrantly expanded CAG repeats that cause HD, a hairpin structure consisting of CAG repeats is formed on one strand and a hairpin structure consisting of CTG repeats is formed on the complementary strand. The CAG hairpin consists of a CAG/CAG unit comprising of an A-A mismatch flanked by C-G base pairs, and the CTG hairpin consists of a CTG/CTG unit with a T-T mismatch sandwiched between C-G base pairs. (Fig. 1b) We were fortunate to discover a CAG repeat-binding molecule, naphthyridine-azaquinolone (NA),[41)] which binds to the CAG/CAG unit.[42)–44)] (Fig. 2a) When this NA was administered to the brain striatum of a mouse model of HD, we found that
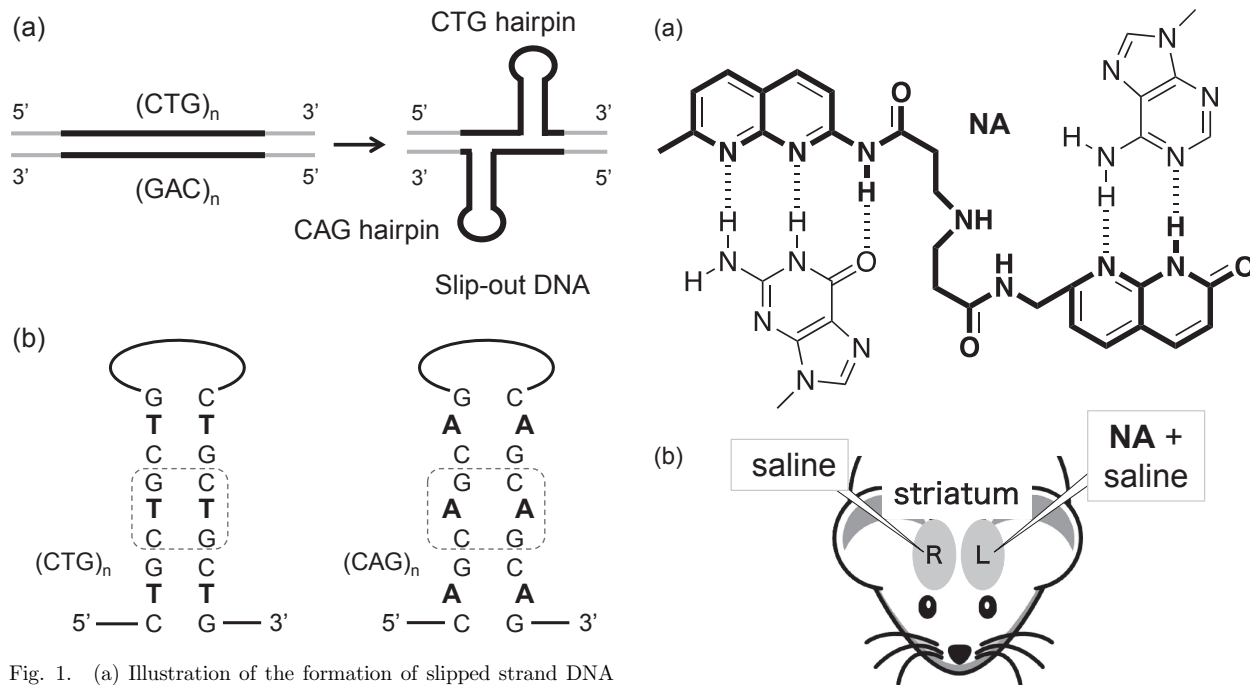
(a)

(b)

Fig. 1. (a) Illustration of the formation of slipped strand DNA (S-DNA) structures in the CTG/CAG repeat. (b) The secondary structures of CTG and CAG hairpins. The dotted line indicated the repeated units of T-T and A-A mismatches flanked by C-G base pairs.

the CAG repeat length was contracted[45] (Figs. 2b and 2c). This suggested the possibility that the administration of small organic compounds may inhibit the onset of the disease by contracting the repeat length, alleviate the symptoms, and even cure the disease by shortening the repeat length below the threshold for onset. In this review, we will first show the *in vivo* effects of our molecules on repeat diseases models, and then explain the binding mode of these molecules to repeat hairpins, and the ideas and concepts from the chemistry viewpoint that led to the development of these molecules, from the last part of the research to the first, as opposed to the flow of the research, to reiterate the real appeal and importance of organic chemistry and synthetic organic chemistry, which deals with molecules at the atomic level. Finally, the challenges and perspective for innovative academic questions derived from this research will be discussed.

Effects of mismatch binding molecules in HD cell and mouse models

The specificity of NA, a molecule that binds to the CAG hairpin structure, for binding to the CAG S-DNA structure was examined by our collaborator Dr. Christopher Pearson (University of Toronto and
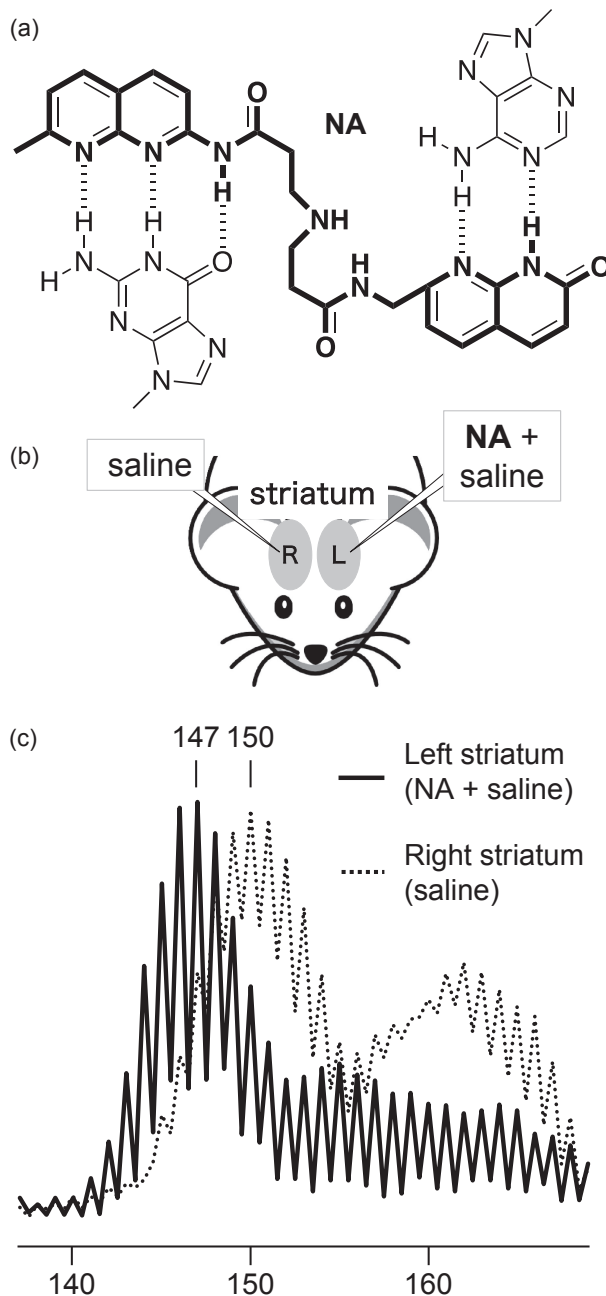
(a)

(b)

(c)

Fig. 2. (a) Chemical structure of CAG repeat-binding molecule **NA** (shown in bold) and illustration of the formation of hydrogen bond pairs with guanine and adenine. (b) Direct injection of **NA** into the striatum of an HD mouse model. (c) Comparison of CAG repeats treated with **NA** (left striatum, solid line) and those treated with saline (right striatum, dotted line) after four injections at 1-week intervals.

Sick Kids Institute, Toronto, Canada) using the CAG/CTG S-DNA with an electrophoretic mobility shift assay (EMSA).[45] As a result, it was confirmed

that NA binds to slip-out with longer CAG repeats and hardly binds to CTG slip-out. *In vitro* repair assays using human cell extracts showed that NA inhibited CAG slip-out repair but not CTG slip-out repair and had no effect on G-T mismatch repair, which plays an important role in DNA damage recovery *in vivo*.[46] To verify the effect of NA on CAG repeat length, Dr. Masayuki Nakamori of Osaka University, a co-investigator of the study, analyzed the expanded CAG repeat length in fibroblasts derived from patients with HD.[47] HD fibroblasts were cultured in the presence and absence of NA for 1 month, and DNA was collected for detailed analysis of the variation in CAG repeat length. The results showed that NA treatment clearly increased the number of cells with shortened CAG repeats, with an average shortening of 9.2 repeats. It was also confirmed that NA had no effect on normal length CAG repeats on other genes. To investigate the effect of NA on the shortening of CAG repeats *in vivo*, we directly administered NA by stereotactic neurosurgery into the striatal tissue of the R6/2 mouse model of HD[48] (Figs. 2b and 2c). We found that the CAG repeat length in the left striatum treated with NA was shorter than that in the right striatum treated with saline. The shortening of CAG repeats in an HD mouse model by NA was also demonstrated to reduce aberrant protein aggregates in the brain striatum, which underlies the pathology of the disease. There was no variation in repeat length in brain tissue not at the site of administration, nor was there any effect on normal length CAG repeats. The mutagenic effect of NA on the entire gene was analyzed using a next-generation sequencer for the *Hprt1* gene, and no significant mutagenic effect was observed. The mechanism of NA-induced shortening of CAG repeats has been investigated, and it has been shown that it involves the formation of CAG slip-outs that occur during DNA transcription, not during DNA replication, and that it promotes the processing of DNA-RNA hybrids (R-loop) during transcription.[49] The reason for contracting the repeat but not expanding the repeat by stabilizing S-DNA structure in this case remains to be clarified.

Spinocerebellar ataxia type 31

SCA31 is a neurodegenerative genetic disease caused by insertions of aberrantly long TGGAA repeat sequences, and the long UGGAA repeats produced by transcription have been reported to exhibit RNA toxicity.[24),25] The pathogenic mechanism of SCA31 is similar to that of trinucleotide repeat diseases, including the formation of RNA foci and the sequestration of RNA-binding proteins (RBPs) in addition to the repeat assisted non-AUG translation.[50)–53] We have started a collaborative study on SCA31 with Dr. Yoshitaka Nagai of Osaka University (currently Kinki University) and Dr. Kinya Ishikawa of Tokyo Medical and Dental University. Because we had no information on small molecules that are likely bind to UGGAA repeats, we screened our own chemical library of repeat-binding molecules using an surface plasmon resonance (SPR) assay with a sensor chip containing immobilized UGGAA repeats. We found naphthyridine carbamate dimer (NCD)[54)–58] as a UGGAA repeat binding molecule[59] (Fig. 3). Details of the screening and binding of UGGAA repeats to NCD will be discussed later.

UGGAA repeats capture RBPs, including splicing factors and form RNA foci in the cell nucleus. It is important to note that the UGGAA repeat sequence is coincidentally found in a long non-coding RNA called HSATIII.[60] HSATIII also sequesters multiple RBPs to form RNA foci. We collaborated with Dr. Tetsuro Hirose of Osaka University to study the effect of NCD binding on the UGGAA repeat RNA-RBP interactions. The results showed that the binding of RBPs, including splicing factors, to the UGGAA repeat RNA was inhibited upon NCD treatment as evaluated using an *in vitro* pull-down assay.[59] In contrast, quinoline carbamate dimer (QCD), a control compound for NCD, did not inhibit RNA-RBP interactions. In addition, the formation of RNA foci from the expression of UGGAA repeats was evaluated by RNA fluorescence *in situ* hybridization; QCD did not affect the formation of RNA foci, but NCD treatment reduced RNA foci in the cell nucleus. In collaboration with Dr. Nagai, we examined the effect of NCDs on RNA toxicity using a Drosophila model of SCA31, which exhibits compound eye degeneration due to the expression of UGGAA repeats in the compound eye. We found that feeding NCD to SCA31 Drosophila model larvae alleviated compound eye degeneration such as reduced compound eye area and lack of pigmentation in adult Drosophila. When QCD was fed as a control compound, the compound did not show any significant effects, suggesting that NCD alleviates RNA toxicity by inhibiting RNA-RBP interactions and RNA foci formation. These results demonstrated that NCD alleviates RNA toxicity caused by UGGAA repeats *in vivo*.
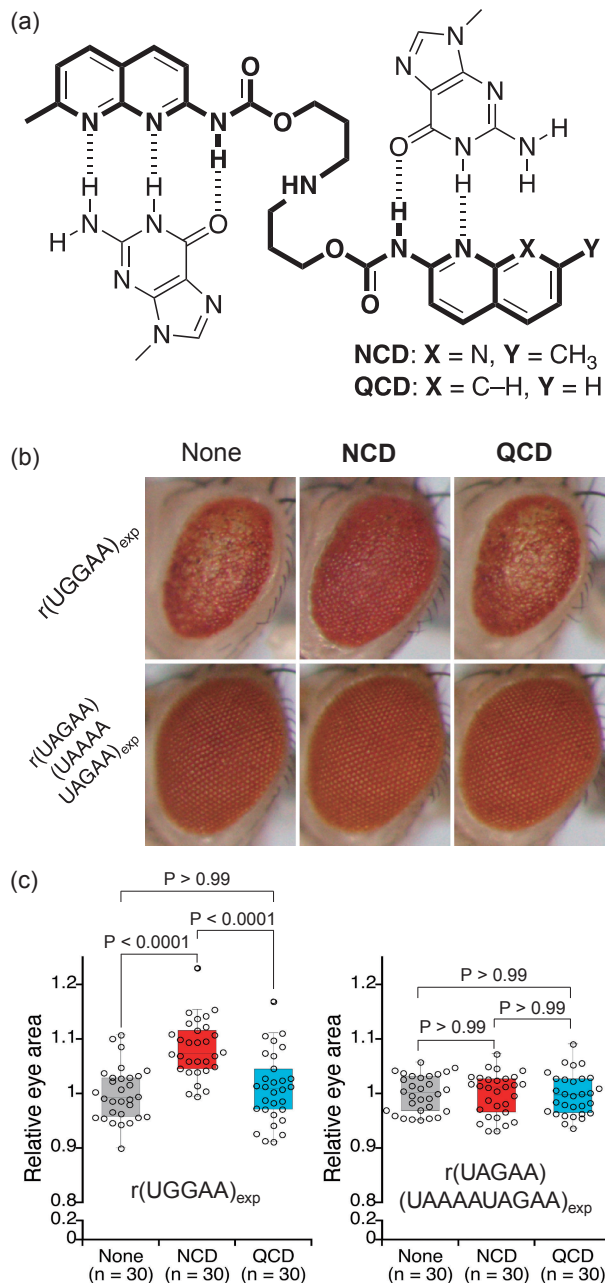
Fig. 3. (a) Chemical structures of **NCD** (X = N, Y = CH₃ shown in bold) and its control compound **QCD** (X = CH, Y = H) and the mode of hydrogen bonding to the two guanines. (b) Alleviation of compound eye degeneration for Drosophila model of SCA31 using **NCD** and **QCD** (top row) and Drosophila expressing non-toxic repeat RNA (bottom row). (c) Quantitative analysis of the effects of **NCD** and **QCD** on the area of compound eyes expressing (left) r(UGGAA)$_{exp}$ and (right) non-toxic repeat RNAs. (Modified from figures in Ref. 59.)

Myotonic dystrophy type 1

In DM1, aberrantly expanded CUG repeats trap mRNA metabolism-related proteins such as MBNL1 and induce selective splicing, which is important for the pathogenesis of the disease.[21)–23)] In our laboratory, we have been working to develop compounds that exhibit binding activity to G-T and G-U mismatches.[61)–64)] Although we have found some effective compounds such as diaminophenanthrene derivatives,[62),63)] their binding selectivity remains an issue because they show affinity for other RNA sequences, such as CCG repeats, in addition to CUG repeats. From structure-activity relationship studies using computer simulations and SPR binding data of synthetic molecules, we found a CUG repeat-selective binding molecule JM642 with a diaminoisoquinoline structure[64)] (Fig. 4). Using a DM1 mouse model, we administered JM642 intraperitoneally (20 mg/kg) for 5 days to investigate the effect of releasing splicing factors trapped by toxic RNA to reduce splicing defects leading to pathogenicity. Only the isoform containing exon 22 (+ex22) is expressed from the *Atp2a1* gene in control WT mice. On the other hand, the expression level of +ex22 isoform was 15.8% in DM1 model mice, but it was greatly restored to 73.6% in JM642-treated mice (Fig. 4b). The reduction in splicing defects was also observed in the *Clcn1* gene (Fig. 4c). Apart from this, RNA aggregate formation, which is frequently observed in myoblasts derived from DM1 patients, was greatly reduced from 41% of 255 cells to 6.7% of 286 cells in JM642-treated cells. These results supported the mechanism of action in which JM642 binds to CUG repeats and inhibits RNA aggregate formation, resulting in the release of trapped splicing factors, suggesting that the selective and high affinity binding properties of JM642 to CUG repeats are important for the splicing aberration recovery effect. By optimizing the substituent at position 5 of 1,3-diaminoisoquinoline, the parent nucleus of JM642, we can expect to create a binding molecule with higher efficacy and lower toxicity.

NA, NCD, and JM642 showed remarkable activity *in vivo*. NA shortened CAG repeats in the striatum of the brain in an HD mouse model,[45)] NCD alleviated the degeneration of compound eyes in a SCA31 Drosophila model,[59)] and JM642 alleviated splicing defects in a DM1 mouse model.[64)] Although more research is needed, all these compounds are expected to be important leads in the development of effective drugs for alleviating symptoms and even
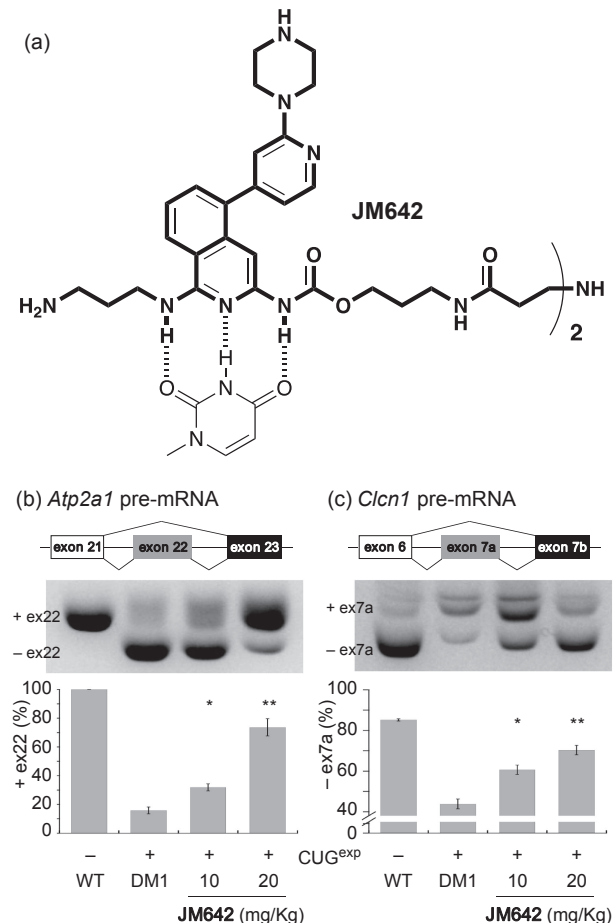
Fig. 4.　(a) Chemical structure of **JM642** (shown in bold) and the proposed hydrogen bonding scheme to uracil. Rescue of missplicing of the pre-mRNAs of the (b) *Atp2a1* and (c) *Clcn1* genes by intraperitoneal injection of **JM642**. Exon 22 is included in the wild-type for the splicing of pre-mRNA of *Atp2a1*, whereas exon 7a is excluded from the pre-mRNA of *Clcn1*. (Modified from figures in Ref. 64.)

inhibiting the onset of neurodegenerative diseases. Each of these compounds has been shown to bind to a target repeat sequence. In other words, compounds that bind to the target repeat sequence are likely to be active *in vivo*, indicating that the evaluation of repeat binding is effective in screening for potential therapeutic agents for repeat diseases. In the following sections, through the use of techniques such as SPR and UV-melting measurements, we will show that these compounds bind to their target repeats.

Screening and analysis by SPR and other methods

In our research, we start by designing and synthesizing molecules that are likely to bind to the target DNAs or RNAs and then evaluate the binding to determine whether to proceed to further biological evaluation.[65]–[68] Therefore, binding evaluation and screening are major milestones in research, and we used SPR and melting temperature measurement of relatively short DNA or RNA as simple and informative methods. In our SPR assay, compounds flowed continuously onto a sensor chip with target DNA or RNA immobilized on a thin gold film via dextran, and the response was recorded. In practice, dielectric constant changes on the thin gold film surface were observed. Because single-stranded repeat DNA or RNA is immobilized on the sensor, it is expected that the repeat DNA or RNA changes into a hairpin structure, *i.e.*, an S-DNA structure, when the compound binds. Upon a structural change from the single-stranded state to the S-DNA hairpin, the distance of the negative charge of the nucleic acid from the gold surface changes significantly, and, therefore, a large response is expected.[42]

Figure 5 shows the results of evaluation with a sensor coated with immobilized $d(CAG)_9$ repeats with different concentrations of NA.[44] In the experiment, the solution containing NA was allowed to flow for 60 seconds from the background condition adjusted with flowing buffer, and only buffer was supplied after 60 seconds. The binding of NA to the CAG repeat DNA on the sensor surface was observed between 0 and 60 seconds, and the dissociation of NA from the NA-CAG repeat complex was observed after 60 seconds. The binding of NA to CAG repeats is characterized by a very slow binding and dissociation response, *i.e.*, slow binding and dissociation rates. This is qualitative, but it can be ratioalized that the single-stranded repeat DNA shows a large structural change.

Another method for the evaluation of binding is change in melting temperature (ΔT_m). This is a technique to observe how the temperature at which a double strand of DNA or RNA melts into a single strand changes depending on whether a test compound is present or not. In double-stranded DNAs and RNAs, the two strands are bound together by hydrogen bonds, and when the temperature is increased, the hydrogen bonds break, and the two strands melt apart to give a single-stranded form. The absorbance at 260 nm increases upon melting of the double strand. Therefore, when the temperature is increased continuously and the absorbance at 260 nm is recorded, the change in absorbance against the temperature gives a sigmoidal graph. The inflection point of this graph is defined as the melting
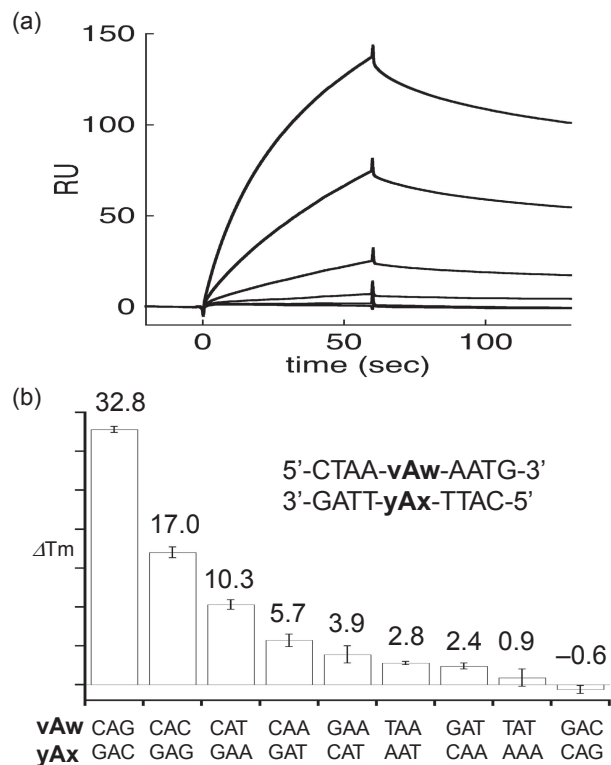
(a)

(b)

5'-CTAA-**vAw**-AATG-3'
3'-GATT-**yAx**-TTAC-5'

vAw	CAG	CAC	CAT	CAA	GAA	TAA	GAT	TAT	GAC
yAx	GAC	GAG	GAA	GAT	CAT	AAT	CAA	AAA	CAG

Fig. 5. (a) SPR analysis of **NA** interaction with the d(CAG)$_9$-immobilized surface. **NA** was added stepwise at 0, 0.13, 0.25, 0.5, 1.0, and 2.0 μM (from bottom to top). The sensors were exposed to the ligand for 0–60 s. d(CAG)$_9$ was immobilized on the streptavidin-coated (SA) sensor chip for 497 RU. (b) Thermal stabilization of A-A mismatches by **NA**-binding. Melting temperature (T$_m$) of 11-mer duplexes containing A-A mismatches in the vAw-xAy sequence in the presence of **NA** is shown above the bar. v-y and x-w denote the Watson-Crick base pairs. Error in T$_m$ measurements was ±0.8 °C and is reported as the standard error of three independent measurements.

temperature (T$_m$) of the duplex. Molecules that bind to DNA or RNA stabilize the double-stranded state, and thus the melting temperature increases. In other words, the degree of binding of a compound can be estimated from the melting temperature that has increased in the presence of the compound. For the same DNA or RNA, the degree of binding can be ranked unambiguously by ΔT$_m$, but for different DNA or RNA, careful evaluation is required, for example, when there is a marked difference in the original T$_m$.

Looking at the actual data, we used double-stranded DNA containing an A-A mismatch (CAG/CAG) sandwiched between C-G base pairs (*cf.* Fig. 1), which appears in the hairpin structure formed by CAG repeats.[42] When NA was added to

this DNA, T$_m$ changed from 18.1 °C to 50.9 °C with an increase of 34.8 °C. When the base pairs flanking the A-A mismatch were changed from C-G base pairs to other base pairs, the increase in T$_m$ was drastically reduced. For CAC/GAG DNA, in which one C-G base pair was replaced by a G-C base pair, the T$_m$ change was almost halved to 17.0 °C (from 20.2 °C to 37.2 °C). The most striking result was the T$_m$ of GAC/GAC DNA where the A-A mismatch was sandwiched between G-C base pairs. The ΔT$_m$ was −0.6 °C (from 22.0 °C to 21.4 °C). Because the measurement error was around ±0.5 °C, it was shown that there was no temperature change due to the addition of NA, *i.e.*, NA did not bind to this sequence. The CAG/CAG sequence, which binds NA, and the GAC/GAC sequence, which does not, have the same number and type of bases, and the only difference is the order of the bases. DNA is a polymer consisting of deoxyribose units with asymmetric carbon atoms and has a right-handed double-helix structure. DNA is a chiral molecule. Therefore, DNA with a CAG/CAG sequence and DNA with a GAC/GAC sequence are completely different compounds, and it is not surprising that the binding of NA is different. This series of DNA binding analyses revealed that NA binds not only to the A-A mismatch, but also to the flanking C-G base pairs. A similar sequence selectivity was observed for NCD, which favorably binds to the CGG/CGG (that is 5'-CGG-3'/5'-CGG-3') sequence DNA but only weakly to the GGC/GGC (5'-GGC-3'/5'-GGC-3').[55] It is important to note that NCD was originally found to bind to CGG/CGG DNA and CGG repeat DNA, as discussed in the next chapter NCD was found to bind to UGGAA/UGGAA RNA and UGGAA repeat RNA, too.

SPR screening of compounds that bind to SCA31 RNA

We screened for compounds that bind to UGGAA repeat RNA using SPR.[59] Because there was no information about compounds that bind to this sequence, we decided to choose from compounds that we have synthesized in our laboratory. If the UGGAA repeats adopted an S-DNA structure, the hairpin structure would have a series of GGA/GGA units (Fig. 6a). Because of the high content of guanine, we thought that a compound that forms complementary hydrogen bonds with guanine might bind to it, and we selected 20 compounds containing NCD.[56],[69] On the other hand, as explained by the sequence selectivity of the NA molecule binding to
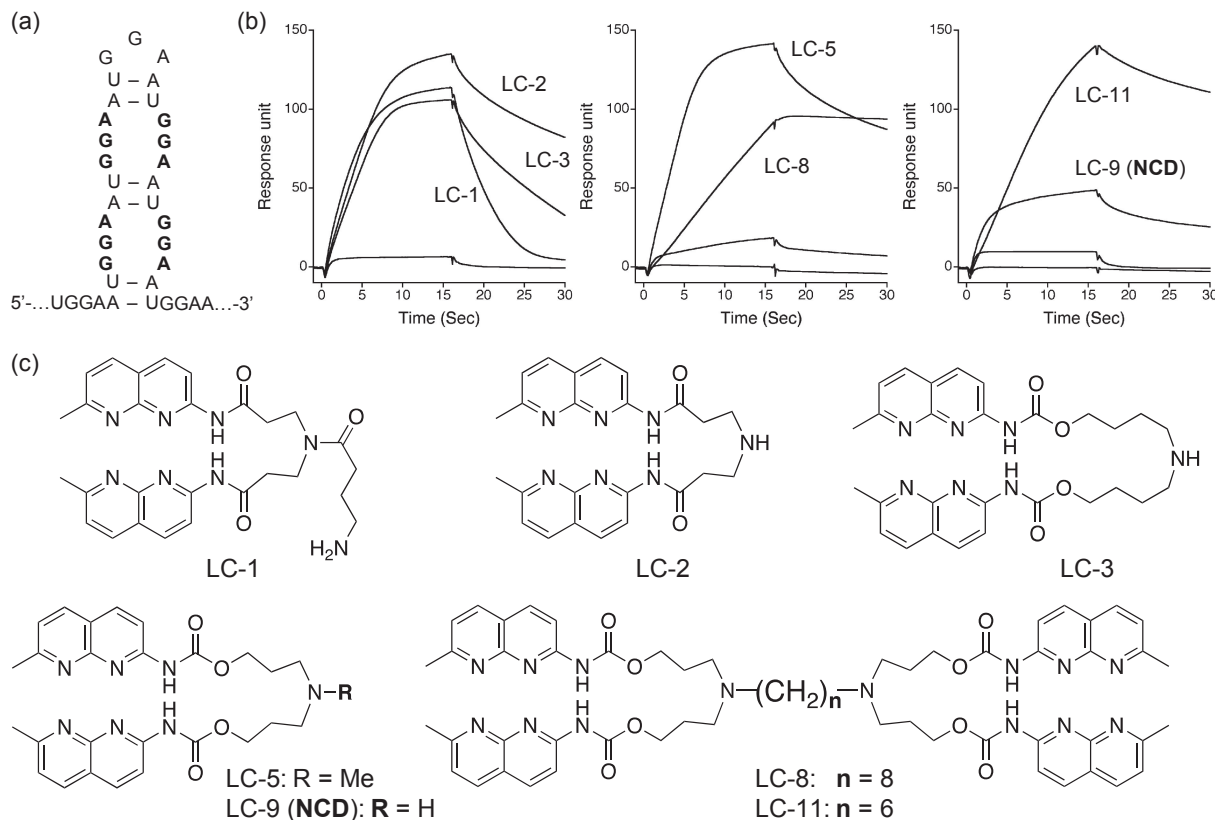
Fig. 6. (a) Proposed secondary structure of UGGAA repeat involving a GGA/GGA internal loop. (b) SPR profiles of compounds applied to the r(UGGAA)$_9$ RNA immobilized on the sensor surface. (c) Structures of compounds exhibited affinity to UGGAA repeat RNA.

CAG repeats, DNA is a chiral molecule; therefore, there is a directionality (5′→3′ or 3′→5′) in the sequence. That is, CGG/CGG (5′-CGG-3′/5′-CGG-3′) DNA to which NCD binds has the 5′-XGG-3′ sequence, but the GGA/GGA (5′-GGA-3′/5′-GGA-3′) repeat unit of the SCA31 RNA has the 5′-GGX-3′ sequence, which is the reverse of the sequence direction. Therefore, we were not confident whether our library of compounds bound to DNA repeats can bind to RNA repeats. We immobilized r(UGGAA)$_9$ on an SPR sensor, evaluated the binding of 20 compounds, and found seven binding compounds. Among the bound compounds, LC-1 and LC-2 were first-generation G-G mismatch binding molecules and their derivatives. LC-3, LC-5, and LC-9 were NCD and their derivatives, and LC-8 and LC-11 were dimers of NCD. These results indicated that LC-9 (NCD) is the smallest structural unit that binds to UGGAA repeat RNA.

The sequence selectivity of NCD in binding to UGGAA repeat RNA was investigated using T_m

measurement. We prepared a control RNA in which guanine in the GGA/GGA sequence was replaced with adenine in a stepwise manner. For these RNAs, NCD was added, and the T_m was measured. When NCD was added to UGGAA/UGGAA double-stranded RNA (T_m 31.1 °C), the T_m increased to 52.6 °C (ΔT_m of 21.5 °C). (Fig. 7a) In the case of the GGA/AGA sequence, in which the 5′-side G of GGA was replaced with A, the ΔT_m by the addition of NCD was markedly reduced to 5.2 °C. Furthermore, in the AGA/AGA sequence in which the 5′-side G of both strands was replaced with A, the effect of NCD was further reduced to 2.7 °C in ΔT_m, and when the G in the center of the sequence was replaced with A, no increase in T_m was observed.

QCD, a control compound of NCD, has a structure in which one of the two 2-amino-1,8-naphthyridine heterocycles of NCD is replaced by a 2-aminoquinoline ring. (Fig. 7b) The methyl group at position 7 was excluded for simplicity of the synthesis. The substitution of a naphthyridine ring with a
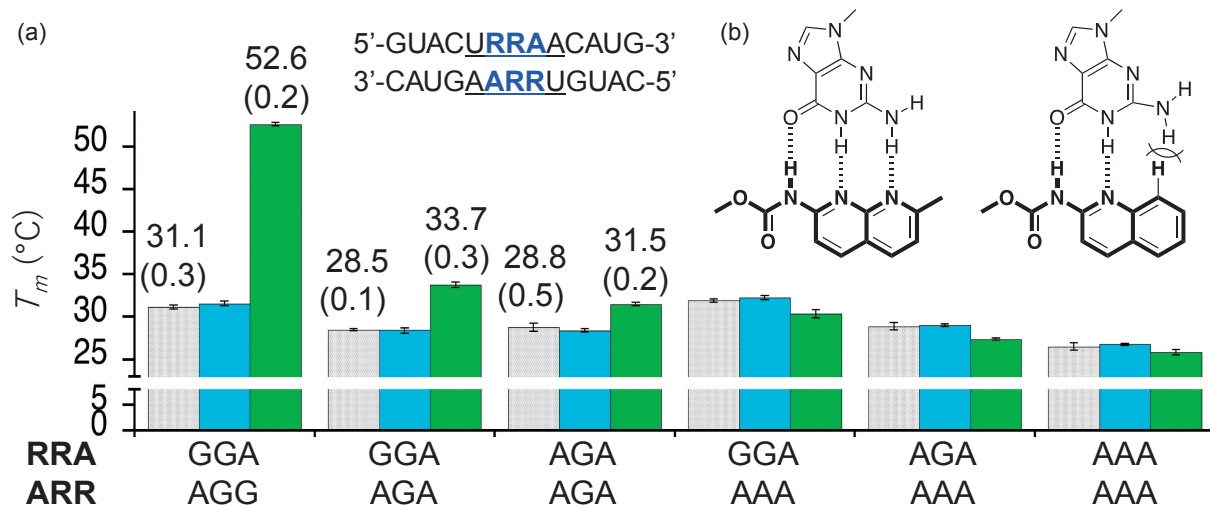
Fig. 7.   (a) Thermal stabilization of RNA duplexes containing URRAA/URRAA pentad in the absence (gray) and presence of NCD (green) and QCD (blue). The data from T_m measurements are presented as the mean ± SD ($n = 3$ independent experiments). (b) The decisive effect of quinoline ring on the hydrogen bonding with guanine.

quinoline ring had a decisive effect on the hydrogen bonding with guanine. The substitution of one nitrogen atom with a carbon (C-H) caused steric repulsion with the amino group hydrogen at position 2 of guanine and prevented hydrogen bonding with guanine. No increase in T_m was observed at all for the UGGAA/UGGAA sequence with QCD, suggesting that the hydrogen bond between NCD and guanine plays a decisive role in the binding of NCD to the UGGAA/UGGAA sequence.

Complex structure of small molecules with repeating DNA and RNA

In our research on the molecular design of organic compounds targeting nucleic acids, it is of course important to know whether they bind or not, but it is of utmost importance to see the complex structure. If we can see the complex structure, we can understand how a small change in functional group or substituent may affect the complex formation based on the structure. Structural information of the complexes is essential for more precise molecular design. Fortunately, in our study, our collaborators were able to analyze the complex structures of NA with CAG/CAG double-stranded DNA[42] and NCD with UGGAA/UGGAA double-stranded RNA[59] using nuclear magnetic resonance (NMR) spectroscopy and complexes of CAG repeat RNA by X-ray analysis.[70]

Structure of the NA-CAG/CAG complex

We asked Dr. Chojiro Kojima (currently a professor at Yokohama National University) to analyze the structure of the NA-CAG/CAG double-stranded DNA complex.[42] The results were clear, and we were able to immediately understand a mystery that we had not understood, namely, why NA binds to the CAG/CAG sequence. NA was designed based on the prediction that each heterocycle consisting of NA, 2-amino-1,8-naphthyridine and 8-azaquinolone, can bind to guanine and adenine, respectively, through hydrogen bonding. In other words, NA is a designed molecule targeting G-A mismatch.[41] However, there is no G-A mismatch in the CAG/CAG sequence, and we did not understand how NA binds to this sequence at all. There was one remaining mystery. From mass spectrometry of the complex, it was known that two molecules of NA bound to the CAG/CAG sequence, but it was not clear why. The structure of the NA-CAG/CAG complex revealed by Dr. Kojima solved both mysteries at once. (Fig. 8a) First, the naphthyridine and azaquinolone rings formed hydrogen bonds with guanine and adenine, respectively. In this respect, the molecular design of NA aimed at binding to the G-A mismatch was found to be successful. The naphthyridine ring deprived the guanine from the hydrogen-bonded C-G base pair. As a result, cytosine, which had no hydrogen-bonding partner,
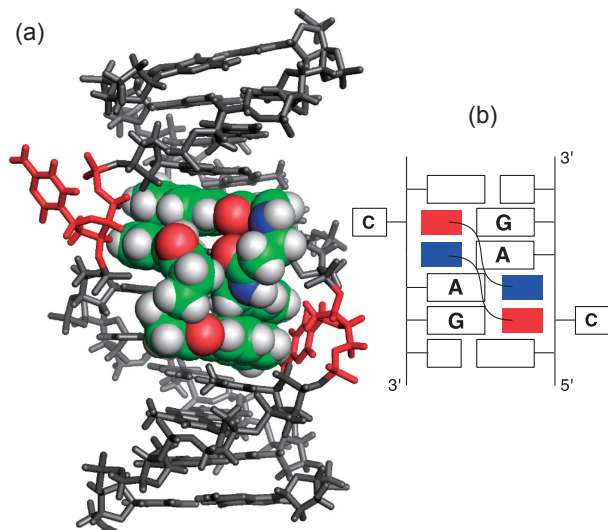
Fig. 8.   (a) NMR-determined structure of the NA-CAG/CAG complex. DNA is shown as a stick model, and the flipped out cytidine residues are shown in red. Two NA molecules are shown in a CPK model. The view is from the major groove. (b) Illustration of a mode of NA-binding to the CAG/CAG site. Rectangular boxes colored in red and blue indicate naphthyridine and azaquinolone heterocycles, respectively. Two cytosines are in a flipped-out position.

was pushed out of the duplex and flipped out. Two molecules of NA were needed to bind the A of one strand of the central A-A mismatch and the G of the other strand. The naphthyridine-G pairs and azaquinolone-A pairs produced well-stacked structure in the interior of DNA (Fig. 8b).

Structure of the NCD-UGGAA/UGGAA RNA complex

The structure of the complex between NCD and UGGAA/UGGAA RNA was analyzed by Dr. Gota Kawai of Chiba Institute of Technology.[59] Mass spectral analysis of the complex indicated that two molecules of NCD bound to UGGAA/UGGAA, and the circular dichroism (CD) spectral measurement also suggested that a large structural change occurred on the RNA side upon binding of NCD. The structure of the complex analyzed is shown in Fig. 9. The naphthyridine ring of NCD formed a hydrogen bond with guanine, which was in good agreement with the importance of guanine indicated by the T_m measurement. Two molecules of NCD bound to the guanine of both strands, as in the NA-CAG/CAG complex.[42] Four guanine-naphthyridine hydrogen bond pairs were stacked well with each other in the interior of the RNA duplex. The
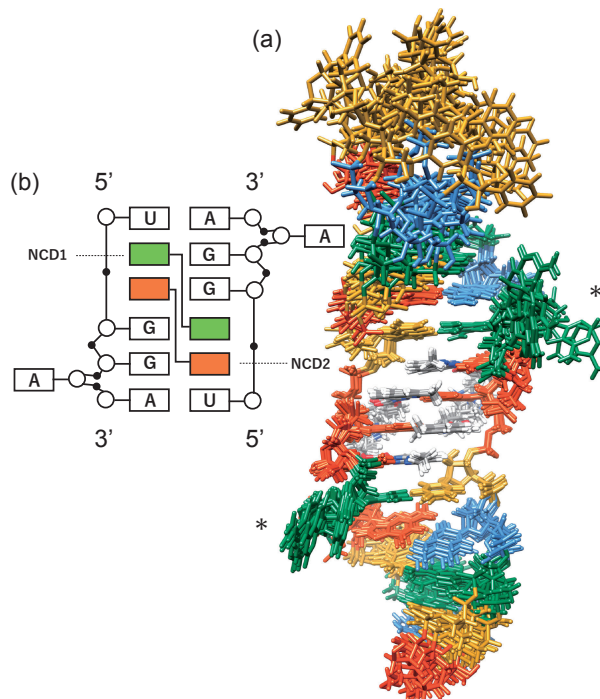

Fig. 9.   (a) NMR-determined structure of the NCD-UGGAA complex viewed toward the major groove side. Nucleotides are colored differently; A (Green), C (sky blue), G (orange), and U (yellow). Two As in the flipped-out position are marked with an asterisk. (b) Illustration of a mode of NCD-binding to the UGGAA/UGGAA site. Rectangular boxes colored in green and orange indicate naphthyridine. Two adenines are in a flipped-out position. (Modified from figures in Ref. 59.)

fourth adenine from the 5′ end of UGGAA/UGGAA was flipped out in the complex (indicated by an asterisk in the figure). The complex structure shown in Fig. 9 is viewed from the major groove side, and the complex of NA-CAG/CAG shown in Fig. 8 is also viewed from the major groove side, but there is a marked difference between the two complex structures. In the NA-CAG/CAG complex structure, the linker connecting the two heterocycles can be seen on the major groove side, whereas it is not the case in the NCD-UGGAA/UGGAA complex structure, indicating that NCD bound from the minor groove side of the UGGAA/UGGAA RNA double strand. The guanine recognition heterocycle, 2-amino-1,8-naphthyridine, formed a hydrogen bond with guanine. Therefore, for the naphthyridine to bind from the major or minor groove side, the order of the hydrogen-bonding groups on the guanine must change accordingly.
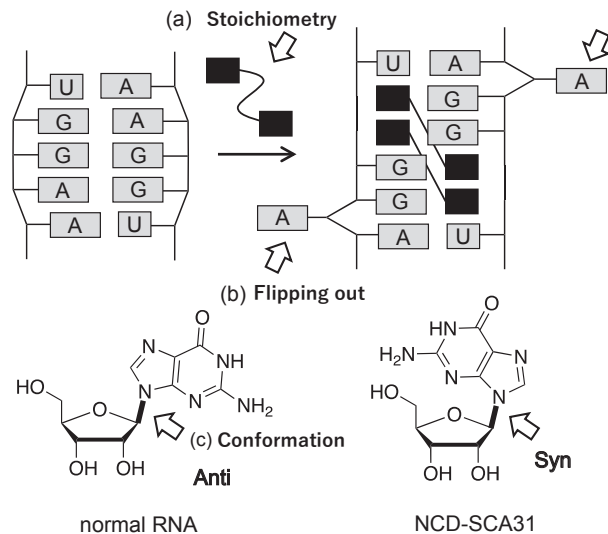
Fig. 10. Characteristic features of NCD-UGGAA/UGGAA RNA binding. (a) Stoichiometry, two NCD molecules bound to the UGGAA/UGGAA site. (b) Flipping out of adenines. (c) Conformation of glycosidic bond.

Features of the NCD-UGGAA/UGGAA RNA complex

Figure 10 shows the structural features of the NCD-UGGAA/UGGAA complex. First, the stoichiometry of the binding is 1 RNA to 2 small molecules, *i.e.*, a binding ratio of 2:1. This is also a characteristic of mismatch binding molecules, as seen in the NA-CAG/CAG and NCD-CGG/CGG complexes. The second feature is the flipping out of nucleobases by binding. In the NCD-UGGAA/UGGAA complex, the fourth adenine base counting from 5′ was flipped out. This base flip out is also a common feature in the binding mode of mismatch-bound molecules.[42),55)] The last feature, which is unique to the NCD-UGGAA/UGGAA structure, concerns the orientation of the glycosidic bond that connects the base and ribose portions of the nucleobase. In normal nucleic acids, the glycosidic bond is anti-type, with the base and sugar portions in an s-trans configuration.[71)] In this case, in guanine, the amino group at position 2, which is the hydrogen bond donor, is located on the minor groove side, and the carbonyl group at position 6, which is the hydrogen bond acceptor, is located on the major groove side. For the formation of a complete hydrogen bond between guanine and 2-amino-1,8-naphthyridine, the amino group at position 2 of the naphthyridine ring must be on the major groove side (see Fig. 3a). In the NCD-UGGAA/UGGAA complex, the linker of the NCD was not on the major groove side, indicating that the glycosidic bond of the guanine to which NCD was bound is not anti-type. A closer look at the NMR structure showed that the glycosidic bonds of the four guanines were of the syn type. When the glycosidic bond was syn-type, the order of hydrogen bonding groups was switched, so the carbonyl group at position 6 was located on the minor groove side and the amino group at position 2 was located on the major groove side. Therefore, there was no question that 2-amino-1,8-naphthyridine binds from the minor groove side. On the other hand, a new question emerged. Does the NCD bind to RNA in which four guanine bases are all syn-type? Or did the configuration of the glycosidic bond change from anti-type to syn-type with the binding of NCD? These questions will be discussed one last time.

Design of mismatch binding molecules

NA, which binds to CAG/CAG DNA, and NCD, which binds to CGG/CGG DNA and UGGAA/UGGAA RNA, discussed so far are compounds that were designed as mismatch binding ligands (MBLs)[65),72)] and showed better-than-expected effects *in vivo*. At the same time, NMR structural analysis of the complexes revealed that they bind to their respective target sequences in a manner different from the designed binding mode. In particular, it has also become clear that factors that are difficult to incorporate in the design are more important. For example, MBLs were designed based on a 1:1 stoichiometry at the design stage, but in reality they formed a 2:1 complex. By flipping out bases that are not involved in the binding, the MBL acquires stabilization through a stacking effect in the interior of duplex. In this section, we will discuss the concept of molecular design of MBLs in detail.

The basis for the molecular design of MBLs was the function of uracil DNA glycosylase (UDG),[73)] which removes uracil bases from DNA, and the mode of its functional expression. Uracil in DNA is mainly caused by mis-incorporation or the deamination of cytosine resulting in the formation of G-U mismatched base pair in duplex DNA (Fig. 11a), and if left unrepaired, adenine is incorporated opposite the uracil during replication, resulting in a G to A mutation. To prevent this mutation, it is necessary to repair the DNA by removing U from the DNA and replacing it with C. The first step in this repair process is the recognition of uracil in the DNA.

The co-crystal structure of UDG and U-containing DNA in the protein data bank (ID 4SKN) shows
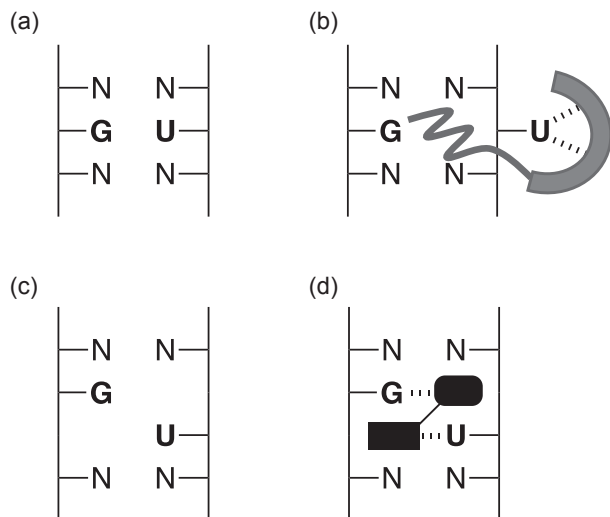
(a) (b)

(c) (d)

Fig. 11. Illustration of the concept of mismatch binding ligand. (a) DNA duplex containing a G-U mismatch. (b) Mode of recognition of U by uracil DNA glycosylase (UDG). (c) Hypothetical structure of G-U mismatch holding space opposite each nucleotide. (d) Binding of two heterocycles (oval and rectangular shapes) in the space opposite G and U.

that UDG inserts an amino acid side chain (Arg272) into the duplex to find a weak G-U base pair by flipping out U and recognizes it in the pocket of UDG (Fig. 11b). We initially planned to reproduce this function in a small molecule, but the mechanism for recognizing the flipped-out U was too complex. Instead, we considered recognizing U in the internal region of the nucleobase stack without flipping out uracil.[65] We considered a hypothetical structure in which the G-U mismatch base pairs are displaced (Fig. 11c), so that neither G nor U has a hydrogen bonding partner and hypothesized that a molecule consisting of a heterocyclic ring that fills the space in front of G and U and forms hydrogen bonds with G and U would be able to stabilize this hypothetical structure.

Molecules that recognize and form hydrogen bonds with nucleobases have been studied for a long time.[74]–[86] A 2-amino-1,8-naphthyridine derivative that forms hydrogen bonds with guanine has been studied.[87]–[89] However, the studies focused exclusively on the interaction in organic solvents and did not evaluate the interaction with DNA in water. When we started our research in the late 1990s, we were inexperienced in molecular recognition research, so we did not know how difficult it was to form hydrogen bond complexes in water.[90] As a result, the competition with water molecules likely had a posi-

tive effect on suppressing the non-specific binding of molecules, because the interior of the nucleic acid is a hydrophobic space. The basic idea behind the MBL design was as follows (Fig. 11d): (1) For the hypothetical structure of a target mismatched base pair, a heterocyclic compound with hydrogen-bonding groups complementary to each nucleobase can be designed. (2) The two heterocyclic compounds are connected by a linker in order to favor the binding entropy. (3) The hydrogen-bonding pair between the heterocyclic ring and the nucleobase is stabilized by stacking in the nucleic acid duplex.[65]

The other important point in MBL design is the charge that the molecule carries. Nucleic acids exist as polyanions due to the negative charge of the phosphate diester moiety, and thus are clothed with positively charged ions in the vicinity. In addition, the presence of divalent cations such as magnesium ions is important for the expression of functions such as ribozymes.[91],[92] We intentionally introduced an amino group in the center of the linker connecting the two heterocycles in the hope that the molecule would carry a positive charge and undergo electrostatic interaction with the nucleic acid, thereby promoting binding to the nucleic acid. This amino group undergoes protonation at neutral pH, bringing a positive charge to the molecule. By extending the linker through this amino group, the MBL can be fixed to a solid phase surface such as an SPR sensor.[41],[66],[93],[94] The length of the linker depends on the estimated degree of freedom of the two heterocycles, but the longer it is, the greater the degree of freedom of the heterocycles and the more entropically unfavorable the binding will be. Because there was no specific index, the length was chosen to be convenient for synthesis and considering the stability of the compound. The first-generation MBL, naphthyridine dimer (LC-2) shown in Fig. 6, which binds to the G-G mismatch sequence in CGG/CGG was reported in 2001.[66],[67] In parallel with this study, we also searched for molecules that bind to bulge structures in DNA double strands.[67] The bulge structure was considered to be part of the hypothetical structure shown in Fig. 11c. Figure 12 shows the binding motif to the cytosine bulge and the structure of the 2,7-diamino-1,8-naphthyridine derivative (DANP) actually bound to the cytosine bulge.[95]–[99] The dimer of 2-amino-1,8-naphthyridine derivative, named ANP77, was found to bind to the stem-loop composed of CC/T and CC/C in DNA.[100],[101] Both DANP and ANP77 have high fluorescence intensity and can be used as fluorescent
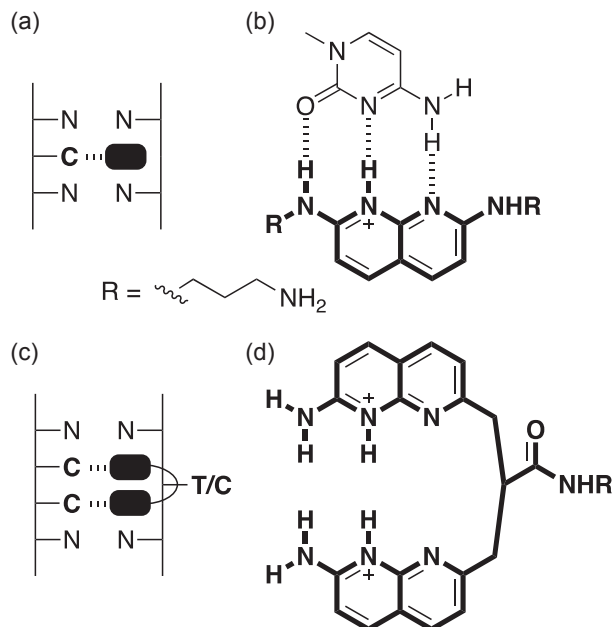
Fig. 12. Illustration of the concept of the bulge binding. (a) Binding of a heterocycle in the space opposite C. (b) Hydrogen bonding scheme of protonated 2,7-diamino-1,8-naphthyridine (DANP) with cytosine. (c) A proposed binding mode of dimeric 2-amino-1,8-naphthyridine derivative (ANP77) to the CC/T and CC/C bulge loops. (d) The structure of ANP77.

probes and indicators to screen out molecules binding to nucleic acids.[102)–105)]

Challenges in designing mismatch-binding ligands

1) Hydrogen-bond pair formation with nucleobases. An important factor in the molecular design of MBLs is the formation of hydrogen-bond pairs with the target nucleobases. So far, we have used 2-amino-1,8-naphthyridine, 8-azaquinolone, and protonated 2-amino-1,8-naphthyridine as heterocycles that form hydrogen-bond pairs with adenine, guanine, and cytosine, respectively. Other heterocycles developed by other researchers are also known.[76),77),82)] On the other hand, the development of heterocycles that bind to thymine (uracil) has not yet been successful in our group. Although JM642 is a characteristic molecule that binds to CUG repeats,[64)] its binding mode has not been clarified. In thymine, the sequence of hydrogen bonding groups is Acceptor (A), Donor (D), and Acceptor (A). It is well known that even if the number of hydrogen bonds increases, the intermolecular interaction does not become stronger in such a sequence of hydrogen bonding groups in which D and A alternate.[106)–108)]

Specifically, 2,6-diaminopyridine, which has a complementary hydrogen-bonding group to thymine, can bind to thymine with three hydrogen bonds, but its binding constant is two orders of magnitude smaller than that of guanine-cytosine, which also has three hydrogen bonds. The design of heterocycles that recognize thymine remains a major challenge.

2) Stoichiometry. Despite the increasing popularity of molecular modeling software and the accuracy of the force field used in the calculations, most MBLs do not bind to the target nucleic acid in the designed binding mode. One of the reasons for this is stoichiometry. Usually, when designing molecules, we assume a 1:1 complex between the target and the molecule, and it is difficult to predict 2:1 complex formation. As already mentioned, MBLs prefer a 2:1 binding ratio instead of 1:1. It may be necessary to develop a molecular design method that incorporates variable binding ratios.

3) Structural change. In the NA-CAG/CAG and NCD-UGGAA/UGGAA complexes, large structural changes occur in both the nucleic acid and the MBL. Such binding might involve induced fit and/or conformational selection mechanisms.[109),110)] Is it possible to predict such large conformational changes in the interaction of small molecules with nucleic acids? The success or failure of the prediction depends on the ability to correctly estimate the stability of the predicted structure. Recently, artificial intelligence that can predict the folding of proteins with high accuracy has been developed and has become a hot topic.[111)] This success was probably due in large part to the abundance of data on protein structures. Unfortunately, data on nucleic acid structures, especially complex structures of nucleic acids and small molecules, are scarce, and the accumulation of data on such complexes will probably be necessary first to be able to predict structural changes involving binding to small molecules.

The intrinsic mystery of MBL binding

In the NA-CAG/CAG complex, there are two naphthyridine-guanine hydrogen-bond pairs, two azaquinolone-adenine hydrogen-bond pairs, and the cytosine that formed the C-G base pair is flipped out. Similarly, in the NCD-UGGAA/UGGAA complex, four naphthyridine-guanine hydrogen bond pairs are formed. Furthermore, in the latter case, all the guanines form syn-type glycosidic bonds. How is this major structural change occurring? We have already mentioned that QCD, in which one of the two naphthyridine rings of NCD is replaced by a

quinoline ring, cannot form a hydrogen bond pair with guanine. However, if QCD bound to the nucleic acid from the naphthyridine ring, would it not be at the last stage of the binding process that the 2-aminoquinoline ring is found to be unable to make hydrogen bond pairs? If this hypothesis is correct, the rate-limiting step in the complex formation between NCD and UGGAA/UGGAA would be in the latter half of the complex formation. Furthermore, does NCD bind only to RNAs with guanine whose glycosidic bonds are in the syn form? Or does the glycosidic bond change from the anti-type to the syn-type during the NCD binding process? In addition, why did NCD not bind to guanine with the anti-type of glycosidic bond from the major groove side? There are many unanswered questions about the binding of MBLs to mismatched base pairs. Although it is unlikely that we will be able to answer all these questions any time soon, the question of how small molecules bind to nucleic acid targets that flexibly change their structure is mysterious and presents a very important scientific challenge.

Expectations for small molecule drug discovery targeting nucleic acids

In recent years, small molecule drug discovery targeting nucleic acids, especially RNA, has shown great promise.[112]–[118] As a result of the encyclopedia of DNA elements (ENCODE; https://www.encodeproject.org)[119] project after the completion of the sequencing of the human genome, it is known that much of the human genome is involved in the maintenance of life as non-coding RNA. If we can find molecules that bind to non-coding RNAs, we can regulate the expression and function of a wide variety of RNAs in the cell using small molecules. The fact that NA-CAG/CAG complex formation shortened CAG repeats in striatal DNA of an HD mouse model and that NCD-UGGAA/UGGAA complex formation alleviated compound eye degeneration in an SCA31 Drosophila model strongly suggests that nucleic acid-targeted small molecule drug discovery is not hypothetical but will be a reality in the near future. In fact, Risdiplam, which was thought to bind to the RNA-protein interface, was approved by the FDA in October 2020 as a treatment for spinal muscular atrophy (SMA).[120] Although antibody drugs and nucleic acid drugs with high drug prices are attracting attention, pharmaceutical companies have accumulated know-how in small molecule drug discovery and a diverse library of organic small molecule compounds. I am convinced that these resources will be quite important and useful for the discovery of small molecules that regulate nucleic acid functions and further development. In order to strongly promote nucleic acid-targeted small molecule drug discovery, it is necessary to gain a deeper chemical understanding of the interactions between nucleic acids and small molecules. At the same time, we need to accelerate our research by making full use of new information science technologies and artificial intelligence.

References

1) Sutherland, G.R. and Richards, R.I. (1995) Simple tandem DNA repeats and human genetic-disease. Proc. Natl. Acad. Sci. U.S.A. **92**, 3636–3641.

2) Moyzis, R.K., Buckingham, J.M., Cram, L.S., Dani, M., Deaven, L.L., Jones, M.D. *et al.* (1988) A highly conserved repetitive DNA-sequence, (TTAGGG)n, present at the telomeres of human-chromosomes. Proc. Natl. Acad. Sci. U.S.A. **85**, 6622–6626.

3) Blackburn, E.H. (1991) Structure and function of telomeres. Nature **350**, 569–573.

4) Harley, C.B., Futcher, A.B. and Greider, C.W. (1990) Telomeres shorten during aging of human fibroblasts. Nature **345**, 458–460.

5) Counter, C.M., Avilion, A.A., Lefeuvre, C.E., Stewart, N.G., Greider, C.W., Harley, C.B. *et al.* (1992) Telomere shortening associated with chromosome instability is arrested in immortal cells which express telomerase activity. EMBO J. **11**, 1921–1929.

6) Mirkin, S.M. (2007) Expandable DNA repeats and human disease. Nature **447**, 932–940.

7) Mitas, M. (1997) Trinucleotide repeats associated with human disease. Nucleic Acids Res. **25**, 2245–2253.

8) Orr, H.T. and Zoghbi, H.Y. (2007) Trinucleotide repeat disorders. Annu. Rev. Neurosci. **30**, 575–621.

9) Verkerk, A.J.M.H., Pieretti, M., Sutcliffe, J.S., Fu, Y.H., Kuhl, D.P.A., Pizzuti, A. *et al.* (1991)

Identification of a gene (*FMR1*) containing a CGG repeat coincident with a breakpoint cluster region exhibiting length variation in Fragile-X syndrome. Cell **65**, 905–914.

10) Fu, Y.H., Kuhl, D.P.A., Pizzuti, A., Pieretti, M., Sutcliffe, J.S., Fu, Y.H. *et al.* (1991) Variation of the CGG repeat at the Fragile-X site results in genetic instability — resolution of the sherman paradox. Cell **67**, 1047–1058.

11) Pieretti, M., Zhang, F.P., Fu, Y.H., Warren, S.T. and Oostra, B.A. (1991) Absence of expression of the FMR-1 gene in Fragile-X syndrome. Cell **66**, 817–822.

12) Knight, S.J.L., Flannery, A.V., Hirst, M.C., Campbell, L. and Christodoulou, Z. (1993) Trinucleotide repeat amplification and hypermethylation of a CpG island in FRAXE mental-retardation. Cell **74**, 127–134.

13) Gecz, J., Gedeon, A.K., Sutherland, G.R. and Mulley, J.C. (1996) Identification of the gene FMR2, associated with FRAXE mental retardation. Nat. Genet. **13**, 105–108.

14) Gu, U. and Nelson, D.L. (2003) FMR2 function: insight from a mouse knockout model. Cytogenet. Genome Res. **100**, 129–139.

15) Macdonald, M.E., Ambrose, C.M., Duyao, M.P., Myers, R.H., Lin, C., Srinidhi, L. *et al.* (1993) A novel gene containing a trinucleotide repeat that is expanded and unstable on huntingtons-disease chromosomes. Cell **72**, 971–983.

16) Duyao, M., Ambrose, C., Myers, R., Novelletto, A., Persichetti, F., Frontali, M. *et al.* (1993) Trinucleotide repeat length instability and age-of-onset in huntingtons-disease. Nat. Genet. **4**, 387–392.

17) Mangiarini, L., Sathasivam, K., Seller, M., Cozens, B., Harper, A., Hetherington, C. *et al.* (1996) Exon 1 of the HD gene with an expanded CAG repeat is sufficient to cause a progressive neurological phenotype in transgenic mice. Cell **87**, 493–506.

18) Taylor, J.P., Hardy, J. and Fischbeck, K.H. (2002) Biomedicine - Toxic proteins in neurodegenerative disease. Science **296**, 1991–1995.

19) Ross, C.A. and Poirier, M.A. (2004) Protein aggregation and neurodegenerative disease. Nat. Med. **10**, S10–S17.

20) Ross, C.A. and Tabrizi, S.J. (2011) Huntington's disease: from molecular pathogenesis to clinical treatment. Lancet Neurol. **10**, 83–98.

21) Brook, J.D., Mccurrach, M.E., Harley, H.G., Buckler, A.J., Church, D., Aburatani, H. *et al.* (1992) Molecular-basis of myotonic-dystrophy — expansion of a trinucleotide (CTG) repeat at the 3′ end of a transcript encoding a protein-kinase family member. Cell **68**, 799–808.

22) Mahadevan, M., Tsilfidis, C., Sabourin, L., Shutler, G., Amemiya, C., Jansen, G. *et al.* (1992) Myotonic-dystrophy mutation — an unstable CTG repeat in the 3′ untranslated region of the gene. Science **255**, 1253–1255.

23) Fu, Y.H., Pizzuti, A., Fenwick, R.G., King, J. and Rajnarayan, S. (1992) An unstable triplet repeat in a gene related to myotonic muscular-dystrophy. Science **255**, 1256–1258.

24) Sato, N., Amino, T., Kobayashi, K., Asakawa, S., Ishiguro, T., Tsunemi, T. *et al.* (2009) Spinocerebellar ataxia type 31 is associated with "inserted" penta-nucleotide repeats containing (TGGAA)n. Am. J. Hum. Genet. **85**, 544–557.

25) Ishiguro, T., Sato, N., Ueyama, M., Fujikake, N., Sellier, C., Kanegami, A. *et al.* (2017) Regulatory role of rna chaperone TDP-43 for RNA misfolding and repeat-associated translation in SCA31. Neuron **94**, 108–124.e7.

26) DeJesus-Hernandez, M., Mackenzie, I.R., Boeve, B.F., Boxer, A.L., Baker, M., Rutherford, N.J. *et al.* (2011) Expanded GGGGCC hexanucleotide repeat in noncoding region of C9ORF72 causes chromosome 9p-linked FTD and ALS. Neuron **72**, 245–256.

27) Renton, A.E., Majounie, E., Waite, A., Simon-Sanchez, J. and Rollinson, S. (2011) A hexanucleotide repeat expansion in C9ORF72 is the cause of chromosome 9p21-linked ALS-FTD. Neuron **72**, 257–268.

28) Mulders, S.A.M., van den Broek, W.J.A.A., Wheeler, T.M., Croes, H.J.E., van Kuik-Romeijn, P., de Kimpe, S.J. *et al.* (2009) Triplet-repeat oligonucleotide-mediated reversal of RNA toxicity in myotonic dystrophy. Proc. Natl. Acad. Sci. U.S.A. **106**, 13915–13920.

29) Wojciechowska, M. and Krzyzosiak, W.J. (2011) Cellular toxicity of expanded RNA repeats: focus on RNA foci. Hum. Mol. Genet. **20**, 3811–3821.

30) Galka-Marciniak, P., Urbanek, M.O. and Krzyzosiak, W.J. (2012) Triplet repeats in transcripts: structural insights into RNA toxicity. Biol. Chem. **393**, 1299–1315.

31) Donnelly, C.J., Zhang, P.W., Pham, J.T., Heusler, A.R., Mistry, N.A., Vidensky, S. *et al.* (2013) RNA toxicity from the ALS/FTD C9ORF72 expansion is mitigated by antisense intervention. Neuron **80**, 415–428.

32) Nalavade, R., Griesche, N., Ryan, D.P., Hildebrand, S. and Krauss, S. (2013) Mechanisms of RNA-induced toxicity in CAG repeat disorders. Cell Death Dis. **4**, e752.

33) Belzil, V.V., Gendron, T.F. and Petrucelli, L. (2013) RNA-mediated toxicity in neurodegenerative disease. Mol. Cell. Neurosci. **56**, 406–419.

34) Jiang, J., Zhu, Q., Gendron, T.F., Saberi, S., McAlonis-Downes, M., Seelman, A. *et al.* (2016) Gain of toxicity from ALS/FTD-linked repeat expansions in C9ORF72 is alleviated by antisense oligonucleotides targeting GGGGCC-containing RNAs. Neuron **90**, 535–550.

35) Marti, E. (2016) RNA toxicity induced by expanded CAG repeats in Huntington's disease. Brain Pathol. **26**, 779–786.

36) Levinson, G. and Gutman, G.A. (1987) Slipped-strand mispairing — a major mechanism for DNA-sequence evolution. Mol. Biol. Evol. **4**, 203–221.

37) Sinden, R.R., Pytlos-Sinden, M.J. and Potaman,

V.N. (2007) Slipped strand DNA structures. Front. Biosci. (Landmark Ed.) **12**, 4788–4799.

38) Wells, R.D. (1996) Molecular basis of genetic instability of triplet repeats. J. Biol. Chem. **271**, 2875–2878.

39) Pearson, C.E., Wang, Y.H., Griffith, J.D. and Sinden, R.R. (1998) Structural analysis of slipped-strand DNA (S-DNA) formed in (CTG)n·(CAG)n repeats from the myotonic dystrophy locus. Nucleic Acids Res. **26**, 816–823.

40) Pearson, C.E., Eichler, E.E., Lorenzetti, D., Kramer, S.F. and Zoghbi, H.Y. (1998) Interruptions in the triplet repeats of SCA1 and FRAXA reduce the propensity and complexity of slipped strand DNA (S-DNA) formation. Biochemistry **37**, 2701–2708.

41) Hagihara, S., Kumasawa, H., Goto, Y., Hayashi, G., Kobori, A., Saito, I. *et al.* (2004) Detection of guanine-adenine mismatches by surface plasmon resonance sensor carrying naphthyridine-azaquinolone hybrid on the surface. Nucleic Acids Res. **32**, 278–286.

42) Nakatani, K., Hagihara, S., Goto, Y., Kobori, A., Hagihara, M., Hayashi, G. *et al.* (2005) Small-molecule ligand induces nucleotide flipping in (CAG)n trinucleotide repeats. Nat. Chem. Biol. **1**, 39–43.

43) Hagihara, M., He, H. and Nakatani, K. (2011) Small molecule modulates hairpin structures in CAG trinucleotide repeats. ChemBioChem **12**, 1686–1689.

44) Li, J., Sakata, A., He, H., Bai, L.-P., Murata, A., Dohno, C. *et al.* (2016) Naphthyridine-benzoazaquinolone: evaluation of tricyclic system for the binding to (CAG)n repeat DNA and RNA. Chem. Asian J. **11**, 1971–1981.

45) Nakamori, M., Panigrahi, G.B., Lanni, S., Gall-Duncan, T., Hayakawa, H., Tanaka, H. *et al.* (2020) Slipped-CAG DNA binding small molecule induces trinucleotide repeat contractions in vivo. Nat. Genet. **52**, 146–159.

46) Pluciennik, A., Burdett, V., Baitinger, C., Iyer, R.R., Shi, K. and Modrich, P. (2013) Extrahelical (CAG)/(CTG) triplet repeat elements support proliferating cell nuclear antigen loading and MutLα endonuclease activation. Proc. Natl. Acad. Sci. U.S.A. **110**, 12277–12282.

47) Sathasivam, K., Amaechi, I., Mangiarini, L. and Bates, G. (1997) Identification of an HD patient with a (CAG)$_{180}$ repeat expansion and the propagation of highly expanded CAG repeats in lambda phage. Hum. Genet. **99**, 692–695.

48) Bates, G.P., Mangiarini, L., Mahal, A. and Davies, S.W. (1997) Transgenic models of Hungtington's disease. Hum. Mol. Genet. **6**, 1633–1637.

49) Salinas-Rios, V., Belotserkovskii, B.P. and Hanawalt, P.C. (2011) DNA slip-outs cause RNA polymerase II arrest in vitro: potential implications for genetic instability. Nucleic Acids Res. **39**, 7444–7454.

50) Zu, T., Gibbens, B., Doty, N.S., Gomes-Pereira, M., Huguet, A., Stone, M.D. *et al.* (2011) Non-ATG-initiated translation directed by microsatellite expansions. Proc. Natl. Acad. Sci. U.S.A. **108**, 260–265.

51) Mori, K., Weng, S.M., Arzberger, T., May, S., Rentzsch, K., Kremmer, E. *et al.* (2013) The C9orf72 GGGGCC repeat is translated into aggregating dipeptide-repeat proteins in FTLD/ALS. Science **339**, 1335–1338.

52) Ash, P.E.A., Bieniek, K.F., Gendron, T.F., Caulfield, T., Lin, W.L., DeJesus-Hernandez, M. *et al.* (2013) Unconventional translation of C9ORF72 GGGGCC expansion generates insoluble polypeptides specific to c9FTD/ALS. Neuron **77**, 639–646.

53) Cleary, J.D. and Ranum, L.P.W. (2013) Repeat-associated non-ATG (RAN) translation in neurological disease. Hum. Mol. Genet. **22**, R45–R51.

54) Peng, T., Murase, T., Goto, Y., Kobori, A. and Nakatani, K. (2005) A new ligand binding to G-G mismatch having improved thermal and alkaline stability. Bioorg. Med. Chem. Lett. **15**, 259–262.

55) Peng, T. and Nakatani, K. (2005) Binding of naphthyridine carbamate dimer to the (CGG)n repeat resulted in the disruption of the G-C base pairing. Angew. Chem. Int. Ed. **44**, 7280–7283.

56) Hong, C., Hagihara, M. and Nakatani, K. (2011) Ligand-Assisted Complex Formation of Two DNA Hairpin Loops. Angew. Chem. Int. Ed. **50**, 4390–4393.

57) Hagihara, M., He, H., Kimura, M. and Nakatani, K. (2012) A small molecule regulates hairpin structures in d(CGG) trinucleotide repeats. Bioorg. Med. Chem. Lett. **22**, 2000–2003.

58) Hong, C., Otabe, T., Matsumoto, S., Dohno, C., Murata, A., Hagihara, M. *et al.* (2014) Formation of ligand-assisted complex of two RNA hairpin loops. Chemistry **20**, 5282–5287.

59) Shibata, T., Nagano, K., Ueyama, M., Ninomiya, K., Hirose, T., Nagai, Y. *et al.* (2021) Small molecule targeting r(UGGAA)n disrupts RNA foci and alleviates disease phenotype in drosophila model. Nat. Commun. **12**, 236.

60) Aly, M.K., Ninomiya, K., Adachi, S., Natsume, T. and Hirose, T. (2019) Two distinct nuclear stress bodies containing different sets of RNA-binding proteins are formed with HSATIII architectural noncoding RNAs upon thermal stress exposure. Biochem. Biophys. Res. Commun. **516**, 419–423.

61) Matsumoto, J., Li, J., Dohno, C. and Nakatani, K. (2016) Synthesis of 1H-pyrrolo[3,2-h]quinoline-8-amine derivatives that target CTG trinucleotide repeats. Bioorg. Med. Chem. Lett. **26**, 3761–3764.

62) Li, J., Matsumoto, J., Bai, L.-P., Murata, A., Dohno, C. and Nakatani, K. (2016) A Ligand that targets CUG trinucleotide repeats. Chemistry **22**, 14881–14889.

63) Li, J., Nakamori, M., Matsumoto, J., Murata, A., Dohno, C., Kiliszek, A. *et al.* (2018) A dimeric 2,9-diamino-1,10-phenanthroline derivative improves alternative splicing in myotonic dystrophy type 1 cell and mouse models. Chemistry **24**, 18115–18122.

64) Matsumoto, J., Nakamori, M., Okamoto, T., Murata, A., Dohno, C. and Nakatani, K. (2020) The dimeric form of 1,3-diaminoisoquinoline derivative rescued the mis-splicing of Atp2a1 and Clcn1 genes in myotonic dystrophy type 1 mouse model. Chemistry **26**, 14305–14309.

65) Nakatani, K. (2009) Recognition of mismatched base pairs in DNA. Bull. Chem. Soc. Jpn. **82**, 1055–1069.

66) Nakatani, K., Sando, S. and Saito, I. (2001) Scanning of guanine-guanine mismatches in DNA by synthetic ligands using surface plasmon resonance assay. Nat. Biotechnol. **19**, 51–55.

67) Nakatani, K., Sando, S., Kumasawa, H., Kikuchi, J. and Saito, I. (2001) Recognition of guanine-guanine mismatch by dimeric form of 2-amino-1,8-naphthyridine. J. Am. Chem. Soc. **123**, 12650–12657.

68) Nakatani, K., Sando, S. and Saito, I. (2000) Recognition of a single guanine bulge by 2-acylamino-1,8-naphthyridine. J. Am. Chem. Soc. **122**, 2172–2177.

69) Dohno, C., Kohyama, I., Hong, C. and Nakatani, K. (2012) Naphthyridine tetramer with a preorganized structure for 1:1 binding to a CGG/CGG sequence. Nucleic Acids Res. **40**, 2771–2781.

70) Mukherjee, S., Błaszczyk, L., Rypniewski, W., Falschlunger, C., Micura, R., Murata, A. *et al.* (2019) Structural insights into synthetic ligands targeting A-A pairs in disease-related CAG RNA repeats. Nucleic Acids Res. **47**, 10906–10913.

71) Saenger, W. (1984) Principles of Nucleic Acids Structure. Springer, New York.

72) Dohno, C. and Nakatani, K. (2011) Control of DNA hybridization by photoswitchable molecular glue. Chem. Soc. Rev. **40**, 5718–5729.

73) Slupphaug, G., Mol, C.D., Kavli, B., Arvai, A.S., Krokan, H.E. and Tainer, J.A. (1996) A nucleotide-flipping mechanism from the structure of human uracil-DNA glycosylase bound to DNA. Nature **384**, 87–92.

74) Hamilton, A.D. and Vanengen, D. (1987) Induced fit in synthetic receptors — nucleotide base recognition by a molecular hinge. J. Am. Chem. Soc. **109**, 5035–5036.

75) Rebek, J., Askew, B., Ballester, P., Buhr, C., Jones, S., Nemeth, D. *et al.* (1987) Molecular recognition: hydrogen bonding and stacking interactions stabilize a model for nucleic acid structure. J. Am. Chem. Soc. **109**, 5033–5035.

76) Zimmerman, S.C. and Weiming, W. (1989) A rigid molecular tweezers with an active site carboxylic acid: exceptionally efficient receptor for adenine in an organic solvent. J. Am. Chem. Soc. **111**, 8054–8055.

77) Zimmerman, S.C., Weiming, W. and Zijian, Z. (1991) Complexation of nucleotide bases by molecular tweezers with active site carboxylic acids: effects of microenvironment. J. Am. Chem. Soc. **113**, 196–201.

78) Park, T.K., Schroeder, J. and Rebek, J. (1991) New molecular complements to imides — complexation of thymine derivatives. J. Am. Chem. Soc. **113**, 5125–5127.

79) Chang, S.K., Vanengen, D., Fan, E. and Hamilton, A.D. (1991) Hydrogen-bonding and molecular recognition — synthetic, complexation, and structural studies on barbiturate binding to an artificial receptor. J. Am. Chem. Soc. **113**, 7640–7645.

80) Shea, K.J., Spivak, D.A. and Sellergren, B. (1993) Polymer complements to nucleotide bases — selective binding of adenine-derivatives to imprinted polymers. J. Am. Chem. Soc. **115**, 3368–3369.

81) Murray, T.J., Zimmerman, S.C. and Kolotuchin, S.V. (1995) Synthesis of heterocyclic-compounds containing 3 contiguous hydrogen-bonding sites in all possible arrangements. Tetrahedron **51**, 635–648.

82) Bell, T.W., Hou, Z., Zimmerman, S.C. and Thiessen, P.A. (1995) Highly effective hydrogen-bonding receptors for guanine derivatives. Angew. Chem. Int. Ed. Engl. **34**, 2163–2165.

83) Beijer, F.H., Sijbesma, R.P., Vekemans, J.A.J.M., Meijer, E.W., Kooijman, H. and Spek, A.L. (1996) Hydrogen-bonded complexes of diaminopyridines and diaminotriazines: Opposite effect of acylation on complex stabilities. J. Org. Chem. **61**, 6371–6380.

84) Fenniri, H., Hosseini, M.W. and Lehn, J.M. (1997) Molecular recognition of NADP(H) and ATP by macrocyclic polyamines bearing acridine groups. Helv. Chim. Acta **80**, 786–803.

85) Inouye, M., Hyodo, Y. and Nakazumi, H. (1999) Nucleobase recognition by artificial receptors possessing a ferrocene skeleton as a novel modular unit for hydrogen bonding and stacking interactions. J. Org. Chem. **64**, 2704–2710.

86) Baudoin, O., Gonnet, F., Teulade-Fichou, M.P., Vigneron, J.P., Tabet, J.C. and Lehn, J.-M. (1999) Molecular recognition of nucleotide pairs by a cyclo-bis-intercaland-type receptor molecule: A spectrophotometric and electrospray mass spectrometry study. Chemistry **5**, 2762–2771.

87) Murray, T.J. and Zimmerman, S.C. (1992) New triply hydrogen bonded complexes with highly variable stabilities. J. Am. Chem. Soc. **114**, 4010–4011.

88) Corbin, P.S. and Zimmerman, S.C. (1998) Self-association without regard to prototropy. A heterocycle that forms extremely stable quadruply hydrogen-bonded dimers. J. Am. Chem. Soc. **120**, 9710–9711.

89) Kelly, T.R., Zhao, C. and Bridger, G.J. (1989) A bisubstrate reaction template. J. Am. Chem. Soc. **111**, 3744–3745.

90) Aoyama, Y. (1991) Molecular recognition of bio-relevant molecules via hydrogen bonding. Nippon Kagaku Kaishi **1991**, 1041–1049 (in Japanese).

91) Piccirilli, J.A., Vyle, J.S., Caruthers, M.H. and Cech, T.R. (1993) Metal-ion catalysis in the tetrahymena ribozyme reaction. Nature **361**, 85–88.

92) Pyle, A.M. (1993) Ribozymes — A distinct class of metalloenzymes. Science **261**, 709–714.

93) Kobori, A., Horie, S., Suda, H., Saito, I. and Nakatani, K. (2004) The SPR sensor detecting the cytosine-cytosine mismatches. J. Am. Chem. Soc. **126**, 557–562.

94) Goto, Y., Suda, H., Kobori, A. and Nakatani, K. (2007) Analysis of mismatched DNA by mismatch binding ligand (MBL)-Sepharose affinity chromatography. Anal. Bioanal. Chem. **388**, 1165–1173.

95) Suda, H., Kobori, A., Zhang, J., Hayashi, G. and Nakatani, K. (2005) N,N′-Bis(3-aminopropyl)-2,7-diamino-1,8-naphthyridine stabilized a single pyrimidine bulge in duplex DNA. Bioorg. Med. Chem. **13**, 4507–4512.

96) Takei, F., Suda, H., Hagihara, M., Zhang, J., Kobori, A. and Nakatani, K. (2007) Allele specific C-bulge probes with one unique fluorescent molecule discriminate the single nucleotide polymorphism in DNA. Chemistry **13**, 4452–4457.

97) Takei, F., Igarashi, M., Hagihara, M., Oka, Y., Soya, Y. and Nakatani, K. (2009) Secondary structure-inducible ligand fluorescence coupled with PCR. Angew. Chem. Int. Ed. **48**, 7822–7824.

98) Takei, F., Igarashi, M., Oka, Y., Koga, Y. and Nakatani, K. (2012) Competitive allele-specific hairpin primer PCR for extremely high allele discrimination in typing of single nucleotide polymorphism. ChemBioChem **13**, 1409–1412.

99) Takei, F. and Nakatani, K. (2017) Fluorescence turn-on hairpin-probe PCR. Chem. Commun. (Camb.) **53**, 1393–1396.

100) Nakatani, K., Natsuhara, N., Mori, Y., Mukherjee, S., Das, B. and Murata, A. (2017) Synthesis of naphthyridine dimers with conformational restriction and the binding to DNA and RNA. Chem. Asian J. **12**, 3077–3087.

101) Das, B., Murata, A. and Nakatani, K. (2021) A small-molecule fluorescence probe ANP77 for sensing RNA internal loop of C, U, and A/CC motifs and their binding molecules. Nucleic Acids Res. **49**, 8462–8470.

102) Zhang, J., Umemoto, S. and Nakatani, K. (2010) Fluorescent indicator-displacement assay for ligand-RNA interactions. J. Am. Chem. Soc. **132**, 3660–3661.

103) Umemoto, S., Im, S., Zhang, J., Hagihara, M., Murata, A., Harada, Y. et al. (2012) Structure-activity studies on the fluorescent indicator in the displacement assay for the screening of small molecules binding to RNA. Chemistry **18**, 9999–10008.

104) Murata, A., Harada, Y., Fukuzumi, T. and Nakatani, K. (2013) Fluorescent indicator displacement assay of ligands targeting 10 microRNA precursors. Bioorg. Med. Chem. **21**, 7101–7106.

105) Fukuzumi, T., Murata, A., Aikawa, H., Harada, Y. and Nakatani, K. (2015) Exploratory study on the RNA-binding structural motifs by library screening targeting pre-miRNA-29a. Chemistry **21**, 16859–16867.

106) Pranata, J., Wierschke, S.G. and Jorgensen, W.L. (1991) OPLS potential functions for nucleotide bases. Relative association constants of hydrogen-bonded base pairs in chloroform. J. Am. Chem. Soc. **113**, 2810–2819.

107) Murray, T.J. and Zimmerman, S.C. (1992) New triply hydrogen bonded complexes with highly variable stabilities. J. Am. Chem. Soc. **114**, 4010–4011.

108) Quinn, J.R., Zimmerman, S.C., Del Bene, J.E. and Shavitt, I. (2007) Does the A·T or G·C base-pair possess enhanced stability? Quantifying the effects of CH···O interactions and secondary interactions on base-pair stability using a phenomenological analysis and ab initio calculations. J. Am. Chem. Soc. **129**, 934–941.

109) Boehr, D.D., Nussinov, R. and Wright, P.E. (2009) The role of dynamic conformational ensembles in biomolecular recognition. Nat. Chem. Biol. **5**, 789–796.

110) Vogt, A.D. and Di Cera, E. (2012) Conformational selection or induced fit? A critical appraisal of the kinetic mechanism. Biochemistry **51**, 5894–5902.

111) Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Tunyasuvunakool, K. et al. (2020) AlphaFold2 Presentation given at CASP 14. https://predictioncenter.org/casp14/doc/presentations/2020_12_01_TS_predictor_AlphaFold2.pdf.

112) Arambula, J.F., Ramisetty, S.R., Baranger, A.M. and Zimmerman, S.C. (2009) A simple ligand that selectively targets CUG trinucleotide repeats and inhibits MBNL protein binding. Proc. Natl. Acad. Sci. U.S.A. **106**, 16068–16073.

113) Jahromi, A.H., Honda, M., Zimmerman, S.C. and Spies, M. (2013) Single-molecule study of the CUG repeat-MBNL1 interaction and its inhibition by small molecules. Nucleic Acids Res. **41**, 6687–6697.

114) Nguyen, L., Luu, L.M., Peng, S.H., Serrano, J.F., Chan, H.Y.E. and Zimmerman, S.C. (2015) Rationally designed small molecules that target both the DNA and RNA causing myotonic dystrophy type 1. J. Am. Chem. Soc. **137**, 14180–14189.

115) Meyer, S.M., Williams, C.C., Akahori, Y., Tanaka, T., Aikawa, H., Tong, Y.Q. et al. (2020) Small molecule recognition of disease-relevant RNA structures. Chem. Soc. Rev. **49**, 7167–7199.

116) Ursu, A., Childs-Disney, J.L., Andrews, R.J., O'Leary, C.A., Meyer, S.M., Angelbello, A.J. et al. (2020) Design of small molecules targeting RNA structure from sequence. Chem. Soc. Rev. **49**, 7252–7270.

117) Falese, J.P., Donlic, A. and Hargrove, A.E. (2021) Targeting RNA with small molecules: from fundamental principles towards the clinic. Chem. Soc. Rev. **50**, 2224–2243.

118) Murata, A., Nakamori, M. and Nakatani, K. (2019) Modulating RNA secondary and tertiary structures by mismatch binding ligands. Methods **167**, 78–91.

119) The ENCODE Project Consortium (2012) An integrated encyclopedia of DNA elements in the human genome. Nature **489**, 57–74.
120) Ratni, H., Ebeling, M., Baird, J., Bendels, S., Bylund, J., Chen, K.S. *et al.* (2018) Discovery of risdiplam, a selective survival of motor neuron-2 (SMN2) gene splicing modifier for the treatment of spinal muscular atrophy (SMA). J. Med. Chem. **61**, 6501–6517.

Profile

Kazuhiko Nakatani was born in Nara Prefecture in 1959. He graduated from the Department of Chemistry, Faculty of Science, Osaka City University in 1982, and went on to graduate school. In 1985, after completing his first year of the doctoral program, he left the university to study under Professor Gilbert Stork at the Department of Chemistry, Columbia University, U.S.A. Three years later, in 1988, he returned to Japan to complete his doctoral degree at Osaka City University and to become a postdoctoral researcher at the Sagami Chemical Research Center in Sagamihara City, Kanagawa Prefecture. In 1991, he started his academic career as a research associate at the Faculty of Science, Osaka City University. Since his student days, he has continued his research on the synthesis of natural products. In 1993, he was hired as a research associate by Professor Isao Saito in the Department of Synthetic Chemistry and Biological Chemistry, Graduate School of Engineering, Kyoto University, and since then, he has been working on nucleic acid chemistry and the interaction of nucleic acids with small organic molecules. He was promoted to Assistant Professor at Kyoto University in 1997, and then to Professor at Osaka University in 2005, where he established the Regulatory Bioorganic Chemistry Laboratory at the Institute of Scientific and Industrial Research, SANKEN. He served as the Director of SANKEN from August 2015 to March 2018, and since August 2019, he has been serving as the Executive Vice President for Finance and Facilities at Osaka University.