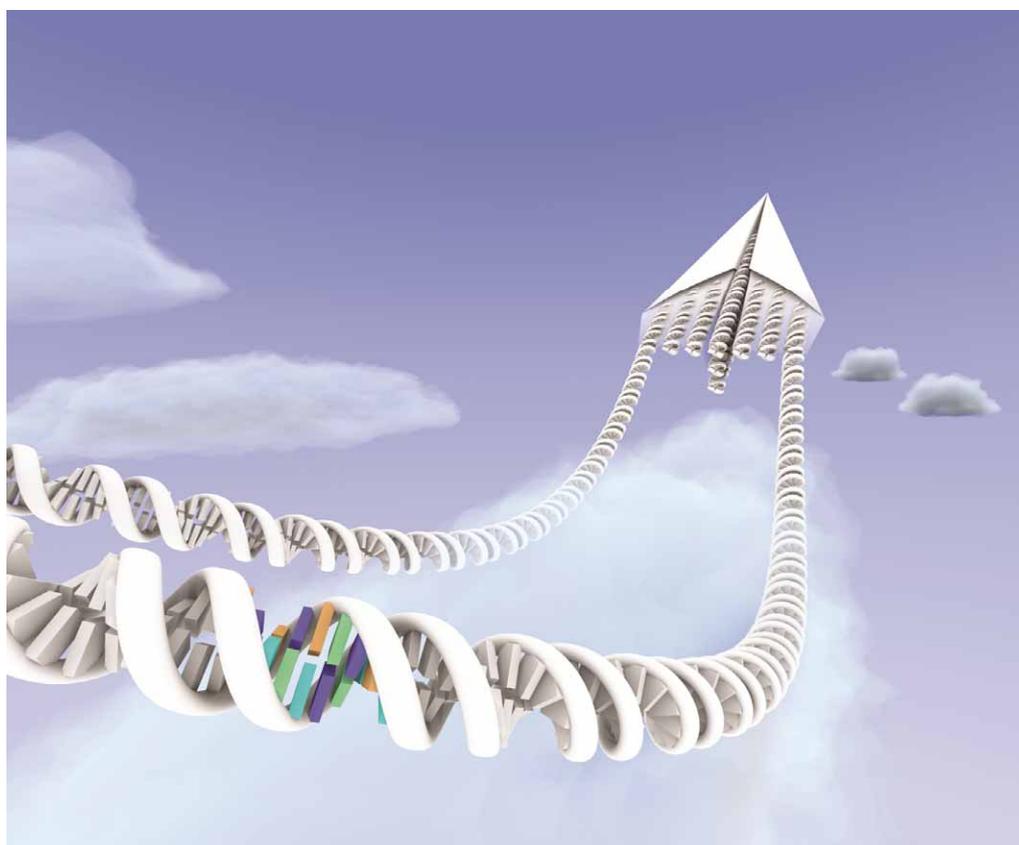


Chem Soc Rev

This article was published as part of the
Advances in DNA-based nanotechnology
themed issue

Guest editors Eugen Stulz, Guido Clever, Mitsuhiro Shionoya and
Chengde Mao

Please take a look at the issue 12 2011 [table of contents](#) to
access other reviews in this themed issue



Cite this: *Chem. Soc. Rev.*, 2011, **40**, 5893–5909

www.rsc.org/csr

CRITICAL REVIEW

Conformational changes of non-B DNA†

Jungkweon Choi and Tetsuro Majima*

Received 2nd June 2011

DOI: 10.1039/c1cs15153c

In contrast to B-DNA that has a right-handed double helical structure with Watson–Crick base pairing under the ordinary physiological conditions, repetitive DNA sequences under certain conditions have the potential to fold into non-B DNA structures such as hairpin, triplex, cruciform, left-handed Z-form, tetraplex, A-motif, *etc.* Since the non-B DNA-forming sequences induce the genetic instability and consequently can cause human diseases, the molecular mechanism for their genetic instability has been extensively investigated. On the contrary, non-B DNA can be widely used for application in biotechnology because many DNA breakage hotspots are mapped in or near the sequences that have the potential to adopt non-B DNA structures. In addition, they are regarded as a fascinating material for the nanotechnology using non-B DNAs because they do not produce any toxic byproducts and are robust enough for the repetitive working cycle. This being the case, an understanding on the mechanism and dynamics of their structural changes is important. In this *critical review*, we describe the latest studies on the conformational dynamics of non-B DNAs, with a focus on G-quadruplex, i-motif, Z-DNA, A-motif, hairpin and triplex (189 references).

1. Introduction

As first proven by Watson and Crick using X-ray diffraction in 1953, DNA has a right-handed helical duplex structure, so-called B-DNA.¹ In living cells, DNA, which acts as the carrier of the genetic information, does not usually exist as a

single-strand sequence, but as a pair of molecules that are held tightly together. However, when the DNA metabolism process occurs, such as replication and transcription, the DNA double helix is partially unwound into a single-strand sequence. Some of the unwound single-strand sequences show repetitive DNA sequences. Generally, it is known that in the human genome, repeat DNA sequences account for more than 50% of the total genomic DNA, whereas simple sequence repeats comprise ~3% of the total DNA.² Under a certain condition, these repetitive DNA sequences can form unique structures rather than double helix of B-DNA. That is, repetitive DNA sequences have the potential to fold into non-B DNA structures such as

The Institute of Scientific and Industrial Research (SANKEN), Osaka University, Mihogaoka 8-1, Ibaraki, Osaka 567-0047, Japan. E-mail: majima@sanken.osaka-u.ac.jp; Fax: +81-6-6879-8495; Tel: +81-6-6879-8499

† Part of a themed issue on the advances in DNA-based nanotechnology.



Jungkweon Choi

Jungkweon Choi received his BS and MS in Chungnam National University in 1992 and 1998, respectively. He received PhD in Chemistry from Kyoto University in 2003. He worked as a research professor in KAIST, Korea, until 2009 and then he is working as a special appointed assistant professor in the Institute of Scientific and Industrial Research (SANKEN), Osaka University, Japan.



Tetsuro Majima

Tetsuro Majima received his BS, MS, and PhD in 1975, 1977, and 1980, respectively. He worked as a research associate in the University of Texas at Dallas for two years (1980–1982), and worked as a scientist in the Institute of Physical and Chemical Research (RIKEN, Japan) for 12 years (1982–1994). In 1994 he became an associate professor in the Institute of Scientific and Industrial Research (SANKEN), Osaka University. He was promoted to a full professor in 1997.

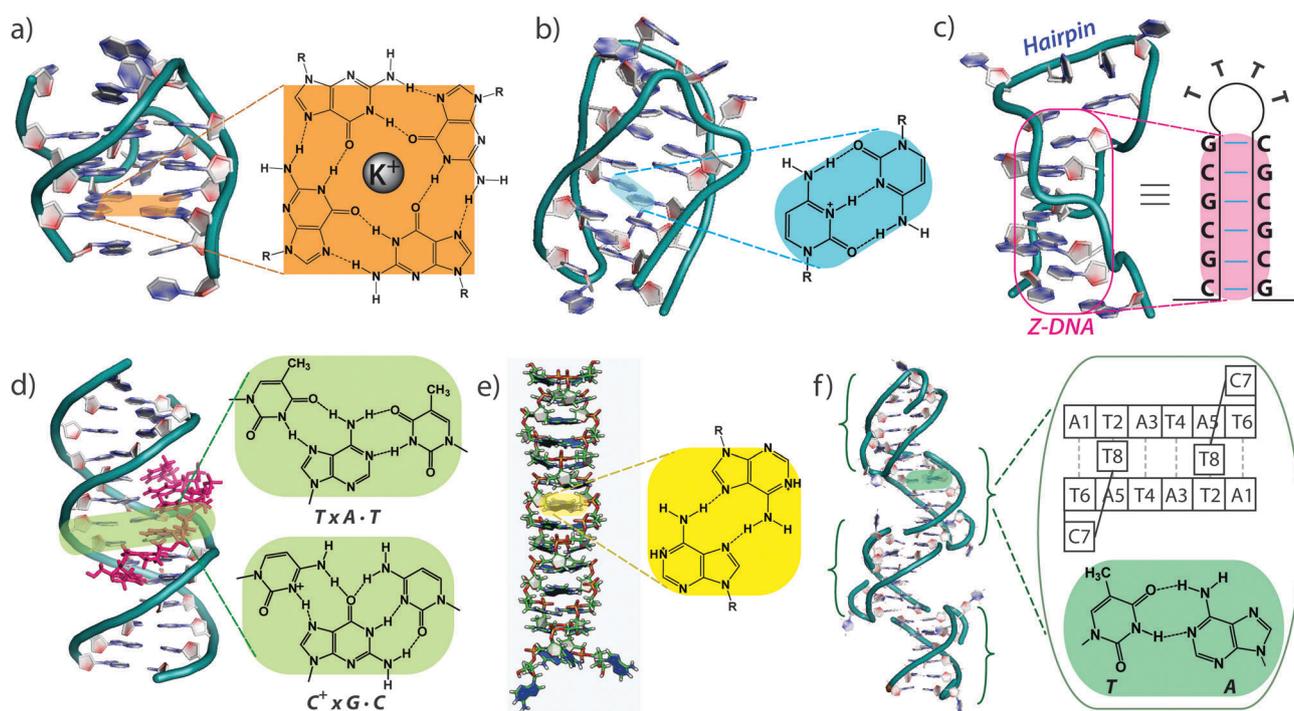


Fig. 1 Representative non-B DNA structures and their molecular structures. (a) G-quadruplex (PDB id: 2KZD) and G-tetrads composed of four guanine bases (orange). (b) i-motif (i-tetraplex) (PDB id: 1EL2) and hemiprotonated C:C⁺ base pair (cyan), (c) hairpin structure with Z-conformation stem shown in pink (PDB id: 1D16), (d) parallel triplex (PDB id: 1BWG) which consists of T × A·T and C⁺ × G·C triads (light green). Triplex forming oligonucleotide (TFO), bound at the major groove of the DNA duplex, is colored in pink. (e) A-motif (taken from ref. 130) and A:A base pair (yellow), and (f) the complex of d(ATATATCT) DNA antiparallel duplex composed of A × T Hoogsteen base pairing fully (PDB id: 2QS6).

hairpin, triplex, cruciform, left-handed Z-form, tetraplex, poly (dA) duplex (A-motif), *etc.* (Fig. 1).^{3–6} These unusual secondary structures may affect the gene metabolism process and also participate in several biologically important processes. According to the previous report by the Vasquez group, the non-B DNA structure-forming sequences can induce genetic instability and consequently may cause human diseases.⁴ In this respect, the molecular mechanism of non-B DNAs has been studied intensively in the pharmaceutical and medical fields. As a result, it was found that non-B DNA induces not only genetic expansions and deletions, but also DNA strand breaks and rearrangements. The non-B DNA structure-induced genetic instability was excellently summarized in recent reviews.^{3–5,7} In addition, many databases have been developed for identifying and evaluating the biological relevance of putative non-B DNA structures in mammalian and other genome.^{8–11} For example, recently, Stephens and coworkers have developed non-B DB, a database to provide access to genome-wide locations of predicted non-B DNA-forming sequence motifs.¹¹ It is freely accessible at <http://nonb.abcc.ncifcrf.gov>.

To date, more than 10 different types of non-B structures have been reported: hairpin, triplex (H-DNA), cruciform, left-handed Z-DNA, tetraplex (G-quadruplex and i-motif), A-motif, sticky DNA, *etc.* In order to form those structures, DNA strands should be folded in a different manner from B-DNA or make unusual base pairs such as Hoogsteen base pairs that are the unusual pairs of hydrogen bonding among nucleic bases compared with Watson–Crick base pairs

(*i.e.*, A–T, G–C). Substantially, Hoogsteen base pairs play an important role in stabilizing several non-B DNA conformations. For instance, in the case of G-quadruplex, i-motif, triplex and A-motif, hydrogen bonds between G–G, C–C, G–G–C and A–A bases, respectively, significantly contribute to their structural stabilization. However, not all of the non-B DNAs have Hoogsteen base pairs; cruciform, Z-DNA and hairpin are stabilized by Watson–Crick base pairs, but the structure is substantially different from B-DNA. It is known that some non-B DNAs such as cruciform and Z-DNA with Watson–Crick base pairs can be induced by outside stimuli such as the negative supercoiling. Additionally, some of the altered conformations with Hoogsteen base pairs are not far different from the canonical B-form.^{12,13} For example, the DNA duplex formed by d(ATATATCT) has two crystal structures: a more stable one with the standard Watson–Crick base pairs and another with Hoogsteen base pairs further stabilized by intercalation of the terminal thymine base (Fig. 1f).¹⁴ Moreover, it has been reported recently that Hoogsteen base pairs are formed transiently even in canonical duplex DNA.¹⁵

On the other hand, DNA is an excellent material for building artificial nanostructures in nanotechnology, material science, molecular computing and bio-analysis because of the Watson–Crick base pairing to make the hybridization between DNA strands highly predictable, the well-defined double-helix structure and its structural stiffness and flexibility.^{16–19} Like B-DNA, non-B DNAs are also regarded as fascinating materials for nanotechnology because they have a unique structure,

do not produce any toxic byproducts and are robust enough for the repetitive working cycle. Moreover, they can be widely used for the applications to biotechnology because many DNA breakage hotspots are mapped in or near the sequences that have the potential to adopt non-B DNA structures.

In this review, we describe the latest studies on the conformational dynamics of non-B DNAs, with a focus on G-quadruplex, C-quadruplex, A-motif and Z-DNA. In addition, we briefly touched on the topics of DNA hairpin and triplex. Since biological functions of nucleic acids significantly depend on their conformational dynamics as well as their structure and stability, an understanding on the mechanism and dynamics of their structural changes is important. Therefore, we believe that this review will be helpful to understand the conformational dynamics of non-B DNAs.

2. G-quadruplex

Guanine (G)-rich sequences can form G-quadruplex structures consisting of π - π stacking of planar G-tetrads, cyclically bound to each other through eight hydrogen bonds according to the Hoogsteen base pairs (Fig. 1a).^{20–22} G-rich sequences are observed frequently in the promoter region of oncogene and human telomeric DNA (Fig. 2).^{23–25} *In vitro*, G-quadruplex blocks the binding of telomerase, and POT1 binding to the single-stranded telomeric DNA enhances telomerase activity by disrupting the G-quadruplex structure, implying the potential biological importance of the G-quadruplex structure in regulating telomere length *in vivo*. In addition, it was recently reported that telomeric DNA can be also subjected to the transcription

process.^{26,27} In this respect, a number of studies on the G-quadruplex have been greatly investigated in terms of the topology, and the molecular and biofunctional mechanism.

To date, various G-quadruplex structures have been characterized by NMR, X-ray crystallography, *etc.* As a result, it is well-known that G-quadruplexes show a high degree of structural polymorphism depending on the nucleotide sequences, the orientation of the strands, the *syn/anti* glycosidic conformation of guanines, the loop connectivities, and environmental factors such as cations, molecular crowding and dehydration (Fig. 2).^{21,22,28–31} In practice, an understanding on the polymorphism of G-quadruplex is very important for determining the role on the populations of its various structure under given conditions and for the application in biotechnology. Although it is known that G-quadruplex structures are stabilized by the presence of monovalent or divalent cations such as Na^+ , K^+ and Pb^{2+} coordinated onto the planar G-tetrads, G-quadruplexes show significantly distinct structures depending on the type of cations.^{32–37} In 1993, the telomere sequence 5'-A(GGGTTA)₃GGG-3' was reported to form an antiparallel G-quadruplex structure in Na^+ solutions.³⁴ However, it was found that the same sequence formed a parallel G-quadruplex structure in K^+ solutions.^{32,33,38,39} On the other hand, Trent and coworkers reported that the 27-mer DNA sequence, d(GGTGGTGGTTGTGGTGGTGGTGG), folds into at least eight different monomeric quadruplex structures under the same experimental conditions.⁴⁰ Moreover, they reported that these different species are produced by rapid cooling, implying significant differences between kinetic and thermodynamic stability. Interestingly, G-quadruplex, which is so-called "a pinched duplex", can be formed in a DNA double helix with consecutive G-G mismatches by the binding of K^+ or Sr^{2+} (Fig. 3a).^{41,42} This unusual pinching conformational change in double-stranded DNA can be used for DNA architecture⁴¹ as well as an electronic nanoswitch.⁴³

DNA G-quadruplex is also revealed to be present as higher-order structures such as dimer and trimer of G-quadruplex subunits.^{30,44,45} For example, Smargiasso *et al.* showed that oligodeoxynucleotides with random bases in the loops, d(GGGW_iGGGW_jGGGW_kGGG) (W = thymine or adenine; *i*, *j*, and *k* = 1–4), form very distinct and stable multimeric G-quadruplexes in the presence of 150 mM K^+ , Na^+ , or NH_4^+ . They also found that sequences with short loops favor a parallel conformation and form very stable multimeric quadruplexes even at low strand concentration, whereas sequences with long loops favor more intramolecular and antiparallel conformations.³⁰ In addition, Lin *et al.* found that addition of K^+ stabilizes the G-quadruplex structure and initial $[\text{K}^+]$ plays a critical role in determining the structural topology of bcl2mid, a kind of bcl-2 gene.⁴⁵ That is, under a low $[\text{K}^+]$ condition, intramolecular G-quadruplex (monomer) is formed as a major component whereas additional intermolecular G-quadruplexes (dimer) are formed at high $[\text{K}^+]$. Furthermore, by adding 145 mM Li^+ , the bcl2mid monomer observed in the 5 mM K^+ condition was converted into the dimer, which was not observed at 150 mM Li^+ solution. Moreover, the T_m of the bcl2mid dimer is higher than that of the monomer by more than 10 °C, indicating the more stable structure of the dimer than that of the monomer.

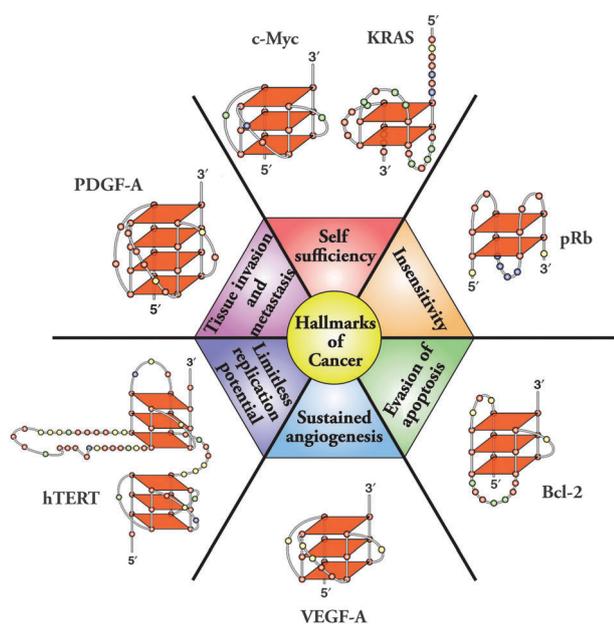


Fig. 2 Six hallmarks of cancer (ref. 25) are shown with the associated G-quadruplexes found in the promoter regions of each oncogene implicated in cancer development and progression. As mentioned in the main text, G-quadruplexes show a high degree of structural polymorphism depending on the nucleotide sequences, the orientation of the strands, the *syn/anti* glycosidic conformation of guanines, the loop connectivities, and environmental factors such as cations and dehydration. This figure was reproduced from ref. 23.

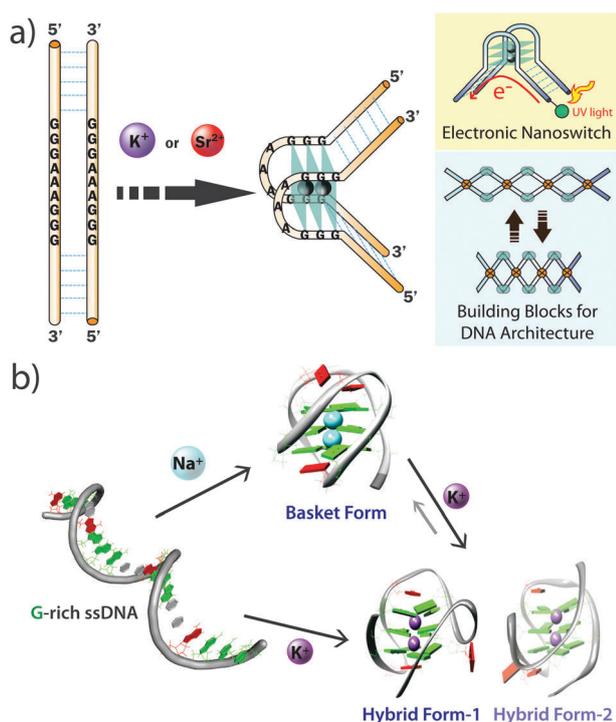


Fig. 3 (a) Illustration of “Pinched DNA” formation triggered by adding cations such as K^+ and Sr^{2+} . The two applications suggested by Sen and coworkers are shown on the right side (reproduced from ref. 41 and 43). (b) Conformational changes of the human telomeric G-quadruplex-forming sequence, d(AGGG(TTAGGG)₃). Two hybrid forms can be induced by adding K^+ , but hybrid-1 structure of which the chain-reversal side loop is near the 5′ end is more stable than hybrid-2 with the side loop near to the 3′ end. This figure is reproduced from ref. 21 and 39.

Additionally, some research groups have attempted to control the G-quadruplex conformation using the chemically modified nucleic acids.^{46–51} In contrast to the ordinary G-rich sequences that show high degrees of structural polymorphism, the incorporation of the chemically modified nucleic acids such LAN (locked nucleic acid) induces a difference structure or a single well-defined structure. For example, the LAN-substituted sequence, GLGLT₄GLGL (L = LAN-G), shows a single well-defined G-quadruplex structure, whereas the native telomeric sequence from *Oxytricha nova*, d(G₄T₄G₄), forms a dimeric G-quadruplex with antiparallel G-columns and diagonal T₄ loops.⁴⁶ The LAN-substituted sequence folds back in a V-shaped turn that puts the first and fourth guanines in the same tetrad, looping over a tetrad with a sharp turn in the DNA backbone, showing how subtle interplay between sequence and conformation defines the folding topology. Saneyoshi *et al.* reported the effect of replacing a single 2′-deoxyguanosine (dG) residue of the 15-base long thrombin-binding aptamer (TBA, 5′-G¹G²T³T⁴G⁵G⁶T⁷G⁸T⁹G¹⁰-G¹¹T¹²T¹³G¹⁴G¹⁵-3′) with methanocarba nucleosides locked in either the C3′-endo (North, N)- or C2′-endo (South, S)-conformation.⁴⁸ They found that the substitution at positions 5, 10 and 14 with a locked South/*syn*-dG nucleoside produced aptamers with the same stability and global structure as the innate, unmodified one, whereas replacing position 15 with the same South/*syn*-dG nucleoside induced a strong destabilization of the aptamer.

Especially, the destabilization induced by the substitution of the South/*syn*-dG nucleoside at position 15 supports the concept that the glycosyl conformation is more restrictive for TBA stability than the sugar pucker. Galeone and coworkers investigated the effect of an 8-methyl-2′-deoxyguanosine residue (M) on the structure and stability of tetramolecular parallel G-quadruplexes, and consequently found that the presence of this residue could result in the formation of unusual quadruplex structures containing all-*syn* tetrads.⁴⁹ Eritja and coworkers reported the effect of 8-amino-2′-deoxyguanosine (8AG) on the stability of G-quadruplex.^{50,51} It was found that 8AG substitution destabilizes the tetraplex structure by 1.2–1.9 kcal mol⁻¹ in the case of the 15-mer long thrombin aptamer (5′-GNTTGGTGTGGTTGG-3′, N = 8AG).

Folding and unfolding kinetics

Since biological functions of biomolecules such as nucleic acids are significantly implicated in their structure and stability as well as their dynamics, studies for the folding and unfolding kinetics of G-quadruplexes can provide new information and a better understanding on their biological function. Indeed, as mentioned above, G-quadruplex-forming sequences have been enormously investigated in terms of their folding topologies and the interactions between small molecules or DNA-binding proteins and G-quadruplex motifs. Despite many studies on G-quadruplex conformational dynamics, however, detailed knowledge of its conformational changes is still needed.

In order to elucidate the folding and unfolding kinetics of both parallel and anti-parallel G-quadruplex observed in K^+ solution, Phan and Patel used the two-repeat human telomeric sequence d(TAGGGTTAGGGT), which can form both parallel and antiparallel G-quadruplex structures in K^+ solution.⁵² They reported that for the modified *UIB7* sequence, d(UAGGGT^{Bt}UAGGGT), the antiparallel G-quadruplex folds faster but unfolds slower than the parallel G-quadruplex at physiological temperature. Hsu *et al.* showed that *c-kit2* (second G-quadruplex-forming motif) exists as an ensemble of structures that share the same parallel-stranded propeller-type conformations.⁵³ They also suggested that using hydrogen–deuterium exchange experiments, at least two structurally similar but dynamically distinct substrates coexist and they undergo slow interconversion on the NMR timescale (> 1 s⁻¹). Adopting the same approach, Balasubramanian and coworkers studied the structure and dynamics of the human telomeric G-quadruplex using single-molecule fluorescence resonance energy transfer (smFRET).⁵⁴ They showed that two stable folded conformations coexist under near-physiological conditions and can interconvert on a minute time scale, and then the unfolding of both substrates occurs at the same rate, which showed dependence on the monovalent metal cation present. Moreover, they also reported that duplex–quadruplex interconversion is a relatively rare event within a natural extended DNA duplex as well as in a single-stranded form.⁵⁵ This is in contrast to the dynamics observed for the human intramolecular quadruplex by Lee *et al.* Lee *et al.* reported that three conformations in vesicle encapsulation studies were observed in K^+ solution, one unfolded and two folded, and each conformation could be further divided into two species,

long-lived and short-lived, based on lifetimes of minutes *versus* seconds.⁵⁶ They showed that folding was severely hindered by replacing a single guanine, showing only the short-lived species and that the long-lived folded states are dominant in physiologically relevant conditions and probably correspond to the parallel and antiparallel G-quadruplexes observed in high-resolution structural studies. The major compaction takes place between the early and late intermediates, and it is possible that local rearrangements are sufficient in locking the late intermediates into the stably folded forms. In addition, telomeric G-quadruplex structures with bromoguanine (BrG)-substitutions were investigated using smFRET.⁵⁷ The observed FRET distributions were composed of five components and the relative population of these components showed the dependence on the position of the BrG-substitution. The results were explained with a structural model, so-called triple-strand-core model. That is, the model includes a triple-stranded core conformation along the folding pathway to hybrid G-quadruplex structures, and equilibrium between hybrid- and chair-quadruplex forms.

Shim *et al.* investigated the folding and unfolding of G-quadruplex aptamer (thrombin-binding aptamer), d(GG-TTGGTGTGGTTGG), in a nanopore nanocavity at the single-molecule level.⁵⁸ As a result, the G-quadruplex formation is found to be cation-selective. The selectivity sequence is $K^+ > NH_4^+ \sim Ba^{2+} > Cs^+ \sim Na^+ > Li^+$, and G-quadruplex was not detected in Mg^{2+} and Ca^{2+} . The high formation capability of the K^+ -induced G-quadruplex is largely due to the slow unfolding reaction ($k_U = 0.066 s^{-1}$). In addition, The Na^+ -quadruplex folds and unfolds most rapidly ($k_F = 6.5 s^{-1}$, $k_U = 2.9 s^{-1}$), while the Li^+ -quadruplex performs both reactions at the slowest rates ($k_F = 0.095 s^{-1}$, $k_U = 0.065 s^{-1}$). Gray and Chaires studied the cation-induced folding into quadruplex structures for three human telomeric oligonucleotides, d[AGGG(TTAGGG)₃], d[TTGGG-(TTAGGG)₃A] and d[TTGGG(TTAGGG)₃], using the combination of equilibrium titrations and multi-wavelength stopped flow kinetics.³⁸ Oligonucleotide folding in 50 mM KCl at 25 °C consisted of single exponential processes with relaxation times τ of 20–60 ms depending on the sequence, whereas folding in 100 mM NaCl consisted of three exponentials with τ -values of 40–85 ms, 250–950 ms and 1.5–10.5 s. From the cation concentration and temperature dependence on the folding rates, they suggest that folding of G-rich oligonucleotides into quadruplex structures proceeds *via* kinetically significant intermediates, which consist of antiparallel hairpins in rapid equilibrium with less ordered structures. Furthermore, they also reported the thermodynamics and kinetics for the conformational changes of the human telomeric G-quadruplex, d[AGGG(TTAGGG)₃], between the Na^+ -basket form and the K^+ -hybrid form (Fig. 3).³⁹ The energy barrier between two conformations was found to be only 1.4–2.1 kcal mol⁻¹ and the transition takes place with three relaxation times: a rapid phase that was complete in <5 ms followed by two slow phases with relaxation times of 40–50 and 600–800 s at pH 7.0 and 25 °C. Furthermore, the addition of TmPyp4 promotes the transition between the basket form and hybrid form (Fig. 3).

In addition, Mergny and coworkers showed that the substitution of 8-amino guanine (8AG) at position 1 of d(TG₄T)

or d(TG₅T) accelerates and stabilizes tetramolecular G-quadruplex formation.⁵¹ This is in contrast to the destabilization of the intramolecular G-quadruplex due to 8AG insertion by Eritja and coworkers.⁵⁰ They illustrated that the driving force for the stabilization of the quadruplex induced by the G to 8AG change is the gain in nucleobase–ion interaction energy, which originates from the better dipole orientation in 8AG.

On the other hand, theoretical and experimental studies for folding pathways of the human telomeric sequences showed that two (3 + 1) G-quadruplex structures are formed through hairpin and triplex intermediates.^{57,59,60} The overall folding would be facilitated by K^+ association in each step, as it decreases the electrostatic repulsion and consequently the energy barrier.⁵⁹ Bardin and Leroy investigated the formation pathway of tetramolecular G-quadruplex with short sequences, d(TG_{*n*}T) (*n* = 3–6). As a result, they found that the quadruplex formation rates increase with the salt concentration but weakly depend on the nature of the counter ions, and the quadruplex formation proceeds step by step *via* sequential strand association into duplex and triplex intermediates.⁶¹ The sequential folding pathway of G-quadruplex was further supported by Rosu *et al.* using electrospray mass spectrometry.⁶² From the observation of tetramolecular G-quadruplex formation from d(TG₅T) in ammonium acetate, they found that the addition of cations promotes the formation of tetramers and pentamers, which are intermediates in the G-quadruplex formation and gradually converted into tetramers. However, the tetramers in this step do not have the perfectly aligned four strands for the five G-quartets, and the rearrangement of the structure towards the well-ordered G-quadruplex structure is revealed to be extremely slow (not complete after 4 months) at 4 °C. Galeone and coworkers reported that the modified nucleoside M, 8-methylene guanine, at the 5'-end of the sequence accelerated quadruplex formation by 15-fold or more relative to the unmodified oligonucleotide, which makes this nucleobase an attractive replacement for guanine in the context of tetramolecular parallel quadruplexes.⁴⁹

The Spomer group has investigated the conformational change of G-quadruplex using molecular dynamics (MD) simulations.⁶³ Their pioneering work published in 1999 emphasized the crucial role of cations on the stabilization of G-quadruplex. From the analysis of MD simulation, it was found that exchange of cations with the solvent is possible without any deformation of the stem, whereas complete removal of the cations from the G-quadruplex causes an immediate destabilization of the structure on the picosecond time scale, followed by the hydration of the vacant cavities on a scale of 100 ps or less. Inspired by the work of Spomer group, a number of attempts have been made to understand the dynamics of G-quadruplex using MD simulations. Among them, Pagano *et al.* compared the dynamics of RNA and DNA G-quadruplexes made from r(GGAGGUUUGGAGG) and d(GGAGGTTTTGGAGG), respectively, using a multi-scale computational approach combining MD simulations and density functional theory calculation.⁶⁴ They found that the coordination process of Na^+ in RNA and DNA G-quadruplex completed in 2 ns and 32 ns, respectively, indicating the faster coordination of Na^+ in RNA G-quadruplex. In addition, adenine bases are found to play an important role in the

stabilization of RNA G-quadruplex because two adenines coordinate to one of two stacked G-tetrads in the presence of Na^+ , equivalent to the energy gain of $4.0 \text{ kcal mol}^{-1}$.

In conclusion, G-quadruplexes show the very slow folding (association) and unfolding (dissociation) kinetics, depending strongly on the number of G-tetrads in the structure and the presence of cations. Furthermore, some recent reports suggest that the structural interconversion among various G-quadruplex conformations occurs only above its melting temperature or high concentrations of DNA strand because an unfolding of the chain for the interconversion is unlikely to happen at room temperature.^{44,45,65} However, this suggestion is in disagreement with the results of other groups who studied the same phenomenon that occurred at room temperature.^{39,66}

3. i-motif DNA (i-tetraplex)

The i-motif structure is formed from a cytosine (C)-rich strand at slightly acidic pH or even neutral pH (Fig. 4).^{23,67} It is known that C-rich sequences are present in or near the regulatory regions of >40% of all genes, especially in the promoter region of oncogene and human telomeric DNA.^{23,68} Thus, i-motif has been an emerging topic in nucleic acids research because they might act as a signpost of oncogene.^{7,23,24,69–72}

The i-motif structure consists of two parallel-stranded C:C+ hemiprotonated base-paired duplexes that are intercalated in an antiparallel manner (Fig. 1b). The i-motif structures are significantly affected by the number of cytosine bases,⁷³ loop length,⁷⁴ environmental condition,^{75,76} and attached or interacting material with the DNA strands.^{77–79} As mentioned above, some sequences showed stable i-motif structures even at neutral pH. For example, the i-motif structure formed from

$d(5\text{mCCT}_3\text{CCT}_3\text{ACCT}_3\text{CC})$ is stable even at neutral pH.⁸⁰ In that case, the i-motif core is extended by a symmetrical T-T pair at one end and by a Hoogsteen A-T pair at the other. Nonin *et al.* showed that the i-motif can also be formed in a dimer of a DNA strand containing two cytidine stretches and an intermediate linker (e.g. $5\text{mCCT}_3\text{AC}_2$) or, again, by intramolecular folding of a single strand with four cytidine stretches.⁸¹ However, the $d((\text{CCATT})_2\text{CCTTTCC})$ sequence found in human centromeric satellite III showed two intramolecular i-motif structures in a pH-dependent ratio.⁸² The two structures differ in intercalation topology, apparently in relation to adenine protonation.

On the other hand, Jin *et al.* investigated the i-motif structure in various pH conditions by small-angle X-ray techniques.⁸³ As a result, they showed that under mild acidic conditions, the conformation of i-motif DNA is similar to that of the partially unfolded i-motif atomic model rather than the fully folded i-motif atomic model. Thus, i-motif DNA is found to be structurally dynamic over a wide pH range, adopting multiple conformations ranging from the folded i-motif structure to a random coil conformation. These results reported until now indicate that like G-quadruplex, i-motif also shows a high degree of structural polymorphism depending on the number of cytosine bases, loop length, environmental condition, and attached or interacting material with the DNA strands. For example, the sequence $5'\text{-CTTTCC-TACCCTCCCTACCCTAA-3}'$ formed multiple “i-motif-like”, classical i-motif, and single-stranded structures as a function of pH.⁸⁴ The classical i-motif structures are predominant in the pH range 4.2–5.2, whereas “i-motif-like” and single-stranded structures are the most significant species in solution at pH higher and lower than that range, respectively. Moreover, Sugimoto and coworkers reported that triplet repeat sequences, $5'\text{-CGG}(\text{CCT})_n\text{CGG-3}'$ ($n = 4, 6, 8$ and 10), could adopt the i-motif structure at neutral pH by molecular crowding.⁸⁵ Dhakal *et al.* found the coexistence of the partially folded form and i-motif in the C-rich human ILPR oligonucleotides using the laser-tweezers technique (Fig. 5).⁸⁶ They also suggested that the formation of i-motif is decreased by increasing pH, while the partially folded structure with a small fraction is pH-independent (pH = 5.5–7.0: 6.1%). In addition, recently we clearly showed that the partially folded species, which could not be observed by the CD spectra, coexist with the single-stranded structure at neutral pH using the FRET technique in the bulk phases and at the single-molecule level. These results imply that the i-motif-like structure may exist *in vivo*, which is a largely neutral condition, and probably participate in biological processes such as replication, regulation, and transcription (Choi, J. *et al.* unpublished data).

Leroy and coworkers have studied the stability of i-motif with regard to its structural polymorphism.^{87–89} The experimental and theoretical studies of $d(\text{Cn})_4$ ($n = 2$ and 4) showed that these sequences could form two i-motif structures, one with outermost 3' extremities and the other with outermost 5' extremities (called 3'E and 5'E topologies, respectively), and the stability in the two topologies is significantly related to an attractive interaction between sugars, responding to the correct backbone twisting.⁸⁷

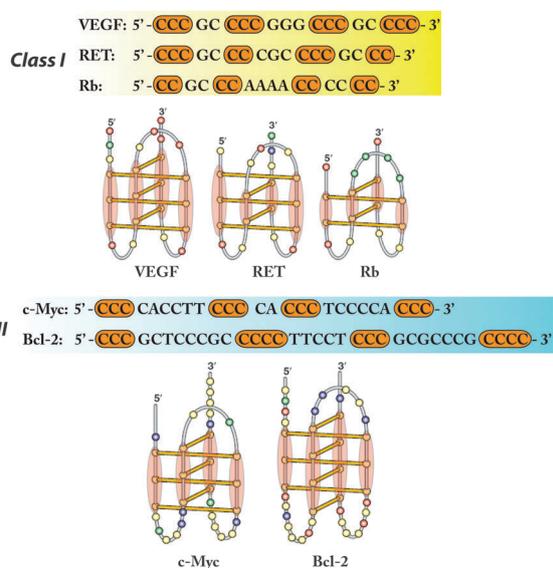


Fig. 4 Sequences and folding patterns of i-motifs in the two proposed classes of i-motifs found in eukaryotic promoter elements. Class I, having small loop sizes, is found in the VEGF, RET and Rb promoter elements, and Class II, having larger loop sizes, is found in the c-Myc and Bel-2 promoter elements. This figure was reproduced from ref. 23 (see ref. 23 for additional details).

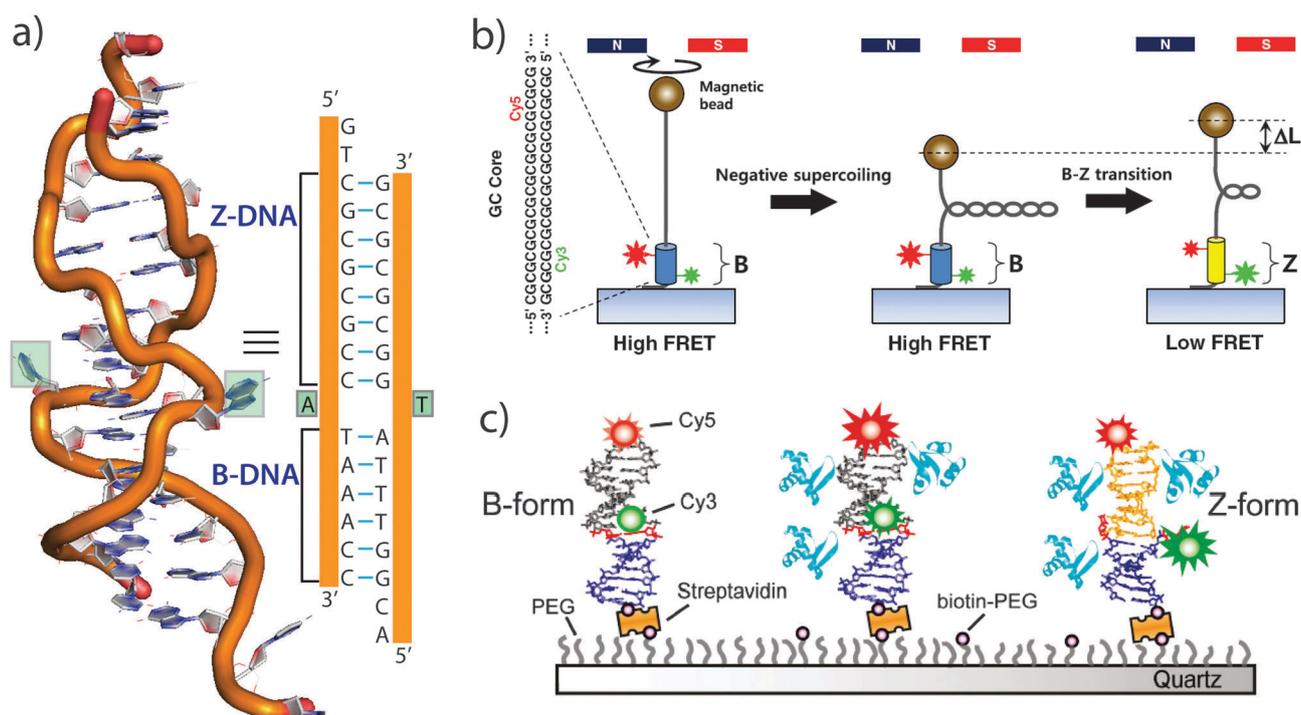


Fig. 7 (a) Crystal structure of the B-Z junction (PDB id: 2ACJ). Adenine and thymine bases at the B-Z junction, flipped out from the axis, are depicted as green boxes. (b-c) Experimental schemes to observe B-Z transition caused (b) by negative supercoiling and (c) Z-DNA binding protein using single-molecule spectroscopy, and details can be found in ref. 119 and ref. 128, respectively.

faster in the 5' → 3' direction than in the 3' → 5' direction, whereas B-DNA showed the faster electron transfer in the 3' → 5' direction.⁹⁶ In addition, they found that the single deep groove of Z-DNA is more hydrated than grooves of B-DNA.⁹⁷

To date, several convincing evidences on the existence of Z-DNA *in vivo* have been suggested. First of all, Z-DNA-forming sequences are located abundantly in the human gene. The Ho group found that Z-DNA-forming sequences exist approximately once every 3000 base pairs by analyzing the complete human genome.⁹⁸ Interestingly, the Z-DNA-forming sequences, especially (GC)_n, are typical sequences of telomeric DNA and a promoter region of many oncogene, suggesting that the Z-DNA formation may be related to carcinogenesis. Ray *et al.* and Wittig *et al.* reported that Z-DNA regulates the oncogene expression such as human ADAM-12⁹⁹ and *c-myc*,¹⁰⁰ respectively. Moreover, various Z-DNA binding proteins (ZBPs) indeed exist *in vivo*, implying that Z-DNA may have a certain role inside the cell.^{101–103} Before the discovery of ZBPs, many researchers were in doubt on Z-DNA, whether it really plays an important role or is just a transient conformation during DNA metabolism. Since the report for the crystal structure of the α domain of ADAR1,¹⁰⁴ which is the human RNA editing enzyme that can bind Z-DNA, Z-DNA regained attention and a number of related studies on Z-DNA and ZBPs have been carried out. As a result, it is known that the binding domain of ZBPs is involved in various biological phenomena from the pathology of viruses¹⁰⁵ to the interaction with rRNA during translation.¹⁰⁶

On the other hand, Ha *et al.* reported the crystal structure of a junction between the right-handed B-DNA and Z-DNA forms of the double helix (Fig. 7a).¹⁰⁷ At the junction, only

one base pair is broken and the bases are flipped out of the duplex. Base pairs of the B-DNA and Z-DNA segments are continuously stacked across the junction. The structure of the B-Z junction maximizes base pairing and stacking, thereby minimizing the energetic cost of the junction and facilitating the use of Z-DNA more widely in nature. Subsequently, Bothe *et al.* showed that the sequence-specific B-DNA flexibility regulates the location of the B-Z junction and thermodynamic property of B-Z transition.¹⁰⁸

B-Z transition

In 1972, Pohl and Jovin showed that the structure of poly(dGC) DNA was converted in 4 M NaCl solution.¹⁰⁹ After several years, Rich and coworkers reported that this conformational change was due to B-Z transition.⁹⁴ The Z-DNA conformation has been difficult to study because it does not exist as a stable feature of the double helix. From numerous experimental and theoretical approaches, it has been known that B-DNA can be converted to Z-DNA by several outside stimuli such as high ionic strength, negative supercoiling, solvent condition, protein binding, and chemical modification.^{105,110,111} As the phosphate groups of Z-DNA are placed closer together than B-DNA, the electrostatic repulsion from positively charged phosphate groups destabilizes the Z-DNA conformation. Under high-salt conditions, however, the electrostatic repulsion is greatly decreased, and consequently Z-DNA conformation is favored.

Recently, Maruyama and coworkers showed that the B-Z transition induced by a cationic graft copolymer, poly(L-lysine)-*graft*-dextran (PLL-*g*-Dex), proceeds in a two-step

manner involving formation of a distinct intermediate.¹¹² In that case, the cationic backbone of the copolymer serves to reduce electrostatic repulsion among DNA phosphate groups and the hydrophilic graft chains reduce activity of water; these two factors act cooperatively to facilitate the B–Z transition. They proposed that the transition rate (8.0×10^{-5} – $3.1 \times 10^{-4} \text{ s}^{-1}$) from intermediate to Z-DNA was the highest when induced by the copolymers with the highest graft contents. Additionally, Tashiro *et al.* showed that B–Z transition can be controlled by changing the temperature; Z-DNA is favored under the low temperature condition.^{113,114}

Second, negative supercoiling is the most plausible Z-DNA inducer inside the cell.^{115,116} Negative supercoiling means twisting the DNA double helix in an unwinding manner (*i.e.*, if DNA is twisted in a winding manner, it is called ‘positive supercoiling’) and it happens during several DNA metabolisms, such as replication and transcription processes.¹¹⁷ To alleviate the torsional stress made by supercoiling, topologically unusual DNA structures such as cruciform and Z-DNA are formed.^{95,118} Recently, Lee *et al.* have examined the B–Z transition induced by the negative supercoiling at the single-molecule level using the combination of FRET and magnetic tweezers (Fig. 7b).¹¹⁹ Magnetic tweezers is a very valuable technique to investigate winding/unwinding processes of coiled DNA molecules by controlling infinitesimal tension and torsion precisely.^{117,120} As a result, they found that B–Z transition can be triggered by minute negative superhelicity ($\sigma \approx -0.006$, superhelical density) and approximately one piconewton tension. In addition, dynamic interconversions between B- and Z-DNA showed the rate constants of the forward ($k_{BZ} = 0.05 \text{ s}^{-1}$) and reverse ($k_{ZB} = 0.07 \text{ s}^{-1}$) reactions at σ of -0.013 and 1.4 pN at 37°C . This result suggests that Z-DNA can be formed *in vivo* more easily, especially in the presence of tension, than expected from the previous studies.^{121,122}

Third, a dehydrating agent such as ethanol, methanol, and ethylene glycol can stabilize the Z-conformation. This stabilization may be due to a closer clustering of counterions around the DNA because ionic effects are felt more strongly, thus providing more effective shielding of the mutually repelling phosphate groups.⁹⁵ Hud and coworkers showed that the $d(\text{GC})_8$ sequence in deep-eutectic solvents (DESS) has a helical structure of Z-DNA, implying non-aqueous solvents can also trigger B–Z transition.¹²³

Fourth, Z-DNA binding protein (ZBP) and antibodies are able to bind Z-DNA (or Z-RNA) selectively, suggesting its Z-conformation triggering ability. Since the method is developed to identify high-affinity Z-DNA binding proteins in 1993,¹²⁴ many researches have been carried out to reveal the binding mechanism upon Z-DNA. In addition, the Qu group reported Alzheimer amyloid protein that can induce Z-to-B transition.¹²⁵ In addition to the previous report on the correlation between Alzheimer’s disease and Z-DNA,^{126,127} this report tentatively suggests that Z-DNA formation may involve the mechanism of Alzheimer’s disease in a certain way. On the other hand, Bae *et al.* investigated the B–Z transition induced by ZBPs to elucidate the detailed binding mechanism and whether the proteins actively induce Z-DNAs or passively trap transiently preformed Z-DNAs (Fig. 7c).¹²⁸ They showed that the

intrinsic B–Z transition rate was considerably similar to that in the presence of ZBPs, suggesting that intrinsic B–Z transition dynamics indeed is present in the relaxed DNA under physiological salt conditions, and ZBPs capture transient Z-DNAs formed naturally by intrinsic B–Z transition dynamics. Therefore, they revealed that ZBPs stabilize Z-DNAs *via* the “conformational selection” mechanism rather than the “induced fit” mechanism.

On the other hand, the chemical modification of bases is one of the factors that stabilize Z-conformation.¹¹⁰ The introduction of the bulky group into a specific base can induce the formation of Z-DNA by the enhanced steric hindrance. Consequently, Z-DNA formed by a chemical modification can exist at physiological salt condition (150 mM NaCl or less). Bromination and methylation at guanine and/or cytosine base are popular ways that take advantage of this principle. Especially, it is known that the substitution of the methyl group at the guanine C8 position stabilizes the Z-DNA conformation of short nucleotides considerably in a variety of sequences.¹²⁹

5. A-motif

Among the repetitive DNA sequences, a single-strand adenine (A)-rich nucleic acid such as polyriboadenylic acid (Poly(A)) and polydeoxyadenylic acid (Poly(dA)) has attracted considerable attention because of its unique structure at acidic pH¹³⁰ and selective binding ability to small molecules.^{131–134} Poly(A) is a tail component of mRNA in all eukaryotic cells and it plays a key role in the stability of mRNA and translation initiation. In eukaryotic mRNAs, the length of the poly(A) tail is initially about 200 As and gradually shortened by exonuclease, an enzyme that degrades mRNA. When a threshold of tail shortening has been reached (about 30 As), the RNA is rapidly destructed.¹³⁵

Poly(dA) (or poly(A)), which is called A-motif, exhibits a single-stranded right-handed helical structure stabilized by the π – π stacking of adenine bases at alkaline and neutral pH, whereas poly(dA) (or poly(A)) at acidic pH forms a right-handed helical duplex with parallel-mannered chains and tilted protonated bases (Fig. 1e and 8). Rich *et al.* proposed that the parallel duplex is stabilized by two factors: the hydrogen bonds (reverse Hoogsteen base-pairing) between two protonated adenine bases and electrostatic attraction between the positively charged protons at the N(1) atom of the adenines and the negatively charged phosphate groups.¹³⁶ However, as originally determined from X-ray data by Rich *et al.*, a full protonation is not a structural requirement for the formation of a parallel-stranded duplex. Several studies for the pH dependent conformational transition of poly(A) proposed that there are three different acidic conformations of poly(A), depending on the extent of protonation of the molecule: A-form, B-form, and frozen form.^{137–139} B-form, also known as an intermediate form, is the only structure present at a pH just below pK_a , associated with the partial protonation of adenine moieties.¹³⁷ As the pH is further lowered, the B-form is gradually converted to the A-form. A-form, which has the “tightly packed” structure, is stabilized by complete protonation of adenine constituents. Finally, the third and most acidic form, known as the “frozen” form, is created from neutral

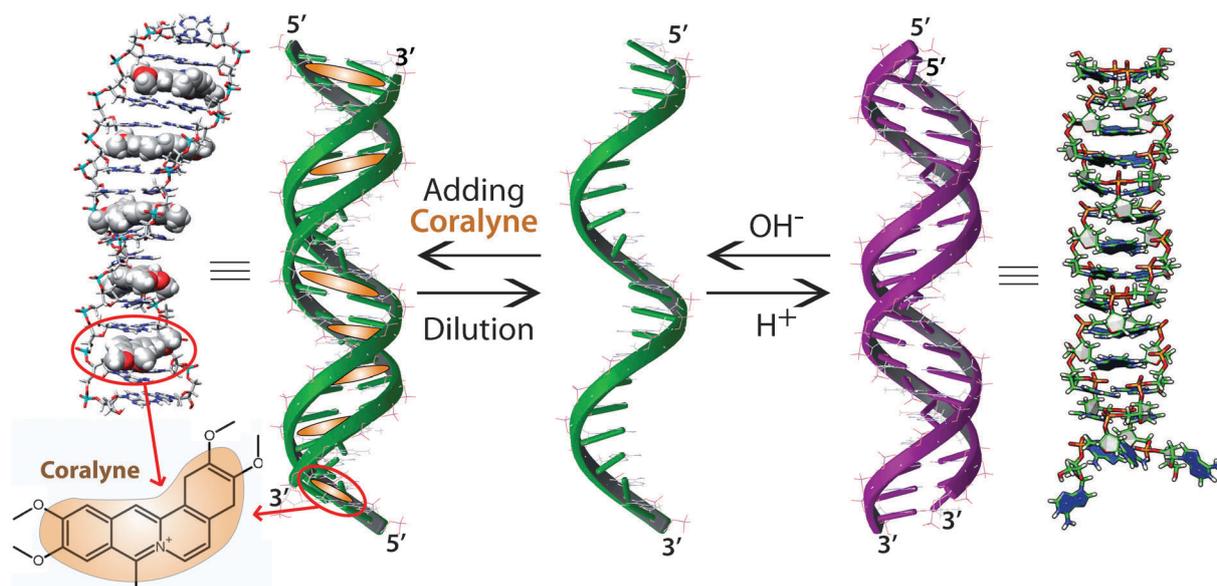


Fig. 8 Conformational change of the A-motif due to pH and the binding of coralyne. Poly(dA) at acidic pH exhibits a right-handed double-stranded helical structure with parallel chains and stacked protonated bases (molecular dynamics (MD) simulation structure from ref. 130 and its schematic duplex (purple), from the right side). However, the binding of coralyne to single-strand Poly(dA) induces the formation of the poly(dA) self-structure, which is an anti-parallel duplex with ostensible A-A base pairs (MD simulation structure taken from ref. 146 and its schematic duplex (green), from the left side).

poly(A) at low pH (<3.8). The frozen form represents a grid-like aggregate consisting of alternating, variably sized, single-stranded regions linked with short double-stranded regions. Petrovic and Polavarapu showed that VCD (vibrational circular dichroism) features are more indicative of the pH-dependent transitions among the three acidic forms.¹⁴⁰ Maggini *et al.* have studied the kinetics of poly(A) double-helix formation and dissociation caused by pH change of the solution.¹³⁸ Using pH-jump and stopped-flow kinetics, they have found that the poly(A) double-helix forms by a second-order step *via* a labile steady-state intermediate and determined the second-order rate constant to be $10^5 \text{ dm}^3 \text{ mol}^{-1} \text{ s}^{-1}$. In the case of the dissociation of the poly(A) duplex, however, it was found to occur much faster ($\sim 20 \text{ s}$) than its formation which continues for 20 minutes or even longer, depending on pH.¹³⁸

Kohler *et al.* investigated the electronic energy relaxation in poly(A) and poly(dA) using the femtosecond transient absorption technique and by steady-state absorption and emission spectroscopies in aqueous solution.¹⁴¹ They showed that the singlet excited states formed in poly(A) and poly(dA) decay on time scales ranging from femtoseconds to nanoseconds, and the secondary structure affects the dynamics of singlet excitations in poly(A) and poly(dA). In addition, the two slowest decay components at room temperature and neutral pH have significantly larger amplitudes in poly(dA) than in poly(A) and double helix formation slightly increases the yield of long-lived excitations. However, the excited-state dynamics of poly(A) and poly(dA) is still not clear until now. Using native PAGE, 2D NMR, circular dichroism (CD) and fluorescence spectroscopy, Krishnan and coworkers characterized the two different pH dependent forms of dA₁₅, single helical structure and parallel-stranded duplex. The time scales of duplex formation

of 5'-TAMRA-dA₁₅ (0.5 μM) was found to be $\tau = 90 \text{ ms}$ demonstrating very fast duplexation, whereas the time scale of duplex dissociation was determined to be $\tau = 7 \text{ s}$.¹³⁰

On the other hand, small crescent-shaped alkaloids such as berberine, palmatine, sanguinarine, and coralyne can bind to poly(dA) (or poly(A)) with high affinity and consequently induce a stable duplex (Fig. 8).^{131,142–146} These alkaloids are representatives of the protoberberine groups with medical importance. Indeed, it is known that coralyne is one of the excellent anti-leukemic drugs and one coralyne molecule is placed in four adenine bases in the poly(dA)–coralyne assembly.¹⁴⁷ The interaction between poly(A) and alkaloids has been enormously investigated in terms of the molecular recognition of poly(A) and the control of a nucleic acid secondary structure. The interaction between poly(A) and alkaloids was excellently summarized in a recent review.¹³¹ In contrast, relatively few studies have been done on poly(dA). In the case of poly(dA), sanguinarine can bind to poly(dA) with the association constant of about $\sim 10^4 \text{ M}^{-1}$.¹⁴⁸ Jain *et al.* showed that coralyne can cause disproportionation of duplex poly(dA)·poly(dT) into triplex poly(dA)·poly(dT)·poly(dT) and the poly(dA) self-structure.¹⁴⁹ Hud and coworkers reported that in the presence of coralyne, the sequence of 3'-(dA)₈-5'-5'-(dA)₈-3', which has different strand polarities, forms the poly(dA)–coralyne assembly with the antiparallel duplex secondary structure from poly(dA) (Fig. 8).¹⁴⁵ Recently, we confirmed that poly(dA) at acidic pH forms a right-handed helical duplex with parallel-mannered chains whereas the coralyne–poly(dA) binding induce a stable antiparallel duplex at neutral pH (Kim, S. *et al.*, unpublished data). Interestingly, Hud and coworkers report that from dilution experiments, the coralyne–poly(A) binding produces very sharp transitions between single-stranded and duplex structures as a function

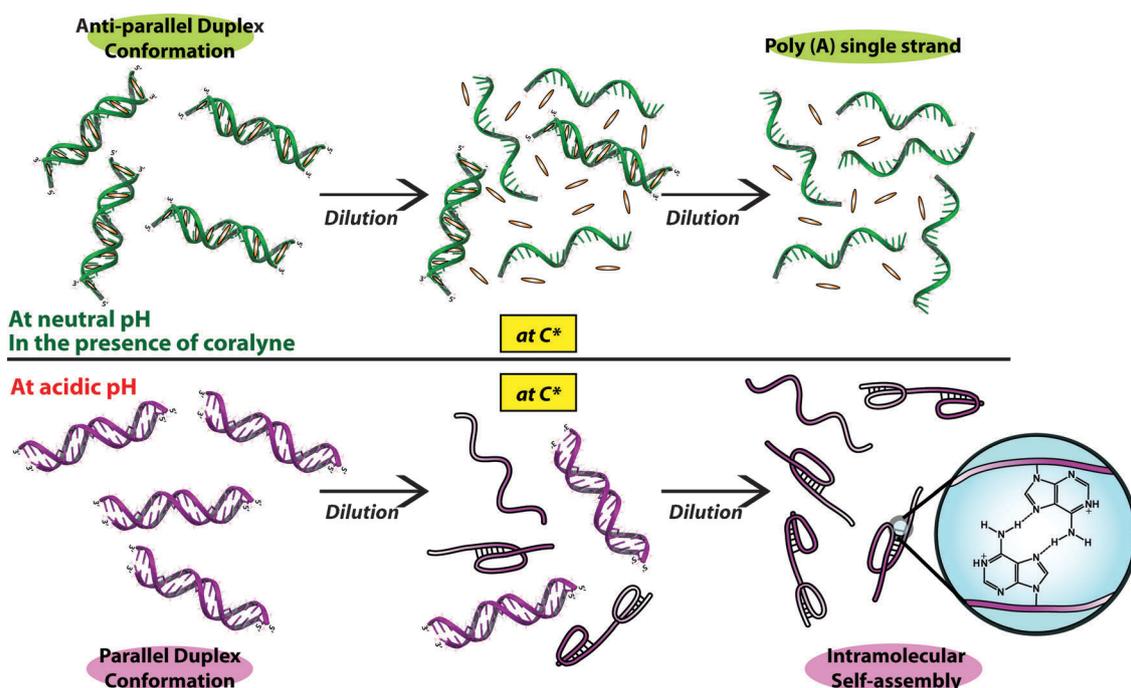


Fig. 9 Dissociation of the anti-parallel duplex of poly(dA) in the presence of coralyne at neutral pH (upper, reproduced from ref. 132) and the parallel duplex of poly(dA) formed at acidic pH (lower). The coralyne–poly(dA) complex begins to be dissociated into the single-stranded structure under a certain concentration (C^*) while the parallel duplex formed at acidic pH is dissociated and converted into other conformations that exhibit intramolecular A:A base pairs.

of ligand concentration, suggesting that the coralyne–poly(A) complex begins to be dissociated under a certain critical concentration (Fig. 9).¹³² However, recently, we found that using the combination of the FRET technique and single-molecule spectroscopy, the parallel duplex formed at acidic pH was converted to another conformation with a smaller hydrodynamic radius, while the coralyne–poly(dA) complex begins to be dissociated into the single-stranded structure under a certain concentration (Fig. 9; Kim, S. *et al.*, unpublished data).

6. DNA hairpin

The hairpin structure might be the most studied and well-understood structural motif of nucleic acids, and its biological function and formation dynamics have been reported by many research groups. For example, in the transcription process, the hairpin structure of mRNA controls the transcription speed depending on the surrounding environment.¹⁵⁰ Moreover, hairpins are building blocks of the secondary structure of RNA and DNA, such as cruciform, pseudoknot, H-DNA (intramolecular DNA triplex) and so on.^{118,151,152} In addition, the hairpin structure is often used for modeling molecular beacon, a new class of nucleic acid probes that make use of sequence-specific interaction of nucleic acid bases.^{153–156} Recently, those probes are investigated for the various purposes in biotechnology, such as the control of gene expression, recognition of a specific nucleic acid sequence, imaging of pathogen RNA in living cells and so on.

Hairpin structure is composed of largely two parts: base paired stem and loop sequence with unpaired or

non-Watson–Crick-paired nucleotides. As discussed later, various factors of the loop and stem sequences such as the type of sequence, loop radius, stem length, *etc.* are closely related to the stability of the hairpin structure. For instance, the GNA (N = A, T, G and C) trinucleotide loop sequence is reported to produce extraordinarily stable DNA hairpins among other possible 64 DNA fragments.¹⁵⁷ In addition, the C:G closing base pair in the stem stabilizes the hairpin with the GNA trinucleotide loop more than G:C or other base pairs.¹⁵⁸ From the MD simulation study, it was found that the stabilization of the C:G closing base pair is due to the additional hydrogen bonding between C of the closing base pair and A of the GNA loop and the good base stacking between the closing base pair and G and A of the loop.

In this section, studies on the folding and unfolding kinetics of DNA hairpin using mainly FRET, FCS and MD simulation techniques are introduced. It is widely known that formation of the hairpin structure occurs in a microsecond time scale, which is much faster than the formation of other non-B DNA structures such as G-quadruplex and i-motif. Using FRET and FCS techniques, Bonnet *et al.* determined the rate constants of the 21-mer DNA hairpin formation and unzipping to be $7.4 \times 10^3 \text{ s}^{-1}$ to $1.4 \times 10^4 \text{ s}^{-1}$ and $1.2 \times 10^2 \text{ s}^{-1}$ to $2.2 \times 10^4 \text{ s}^{-1}$, respectively, in the range of 10 °C to 45 °C (Fig. 10a).¹⁵⁹ Interestingly, the rate of hairpin formation was increased by shorter loop length, higher salt concentration and $d(T)_n$ loop rather than that of $d(A)_n$. Wallace *et al.* also studied the motion of a dye-labeled DNA hairpin loop (Cy5-5'-GGGTT-(A)₃₀-AACCC-3'-TMR) in aqueous solution, Tris-EDTA buffer and Tris-EDTA buffer containing an excess of DNA complementary to the loop sequence, (T)₃₀.¹⁶⁰ From the

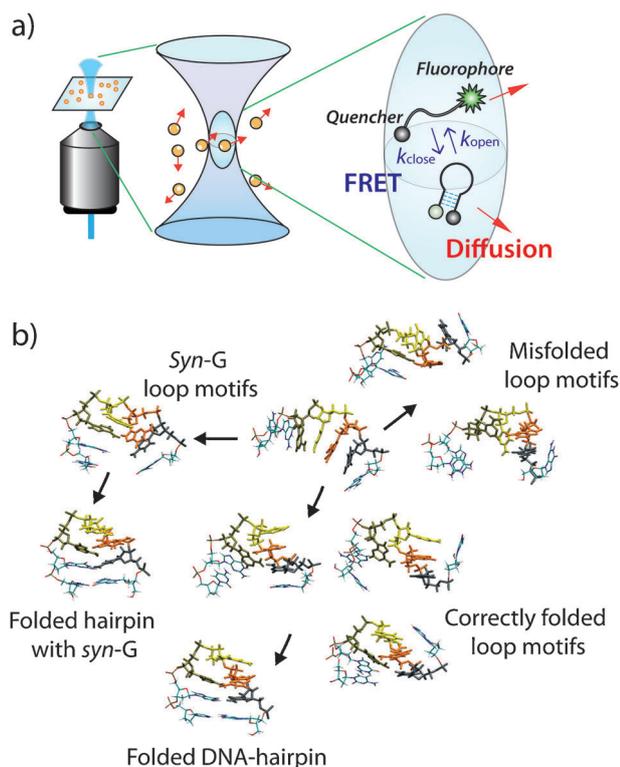


Fig. 10 (a) Representative illustration of the experimental scheme of Bonnet *et al.* (see the details in ref. 159) using FRET and FCS for the study on DNA hairpin formation. (b) Folding intermediates of the DNA-trinucleotide hairpin loop. Each of the snapshots from various stages is from the RexMD simulation (taken from ref. 162).

comparison between obtained histograms and simulated FRET efficiency curves, the hairpin sample in aqueous solution showed wider distribution than the simulated curve, indicating the existence of many conformations due to a lack of stabilizing counterions. However, in buffer solution (Tris-HCl, EDTA or NaCl), the mean correlation time corresponding to the average relaxation time for the open-to-closed transition is reduced, suggesting that the hairpin is now a stable structure and binding in the stem sequence is the most likely fluctuation phenomenon (*i.e.*, the hairpin open to closed transition). Meanwhile, the addition of complementary DNA showed a narrow distribution of FRET efficiency well agreed with simulated curves and its β value, a stretch parameter describing the heterogeneity of the system, was the highest among three samples. This means that the major contribution to the fluctuation in the case of the third sample is motion of the stem arms and not the open-to-closed transition of the hairpin.

On the other hand, Zewail and coworkers studied folding and melting dynamics of the 25-mer DNA hairpin (5'-R6G-CCCCTTAGTAGTTCCTCACAAGGGG-3') using an ultrafast T-jump (<20 ps).¹⁶¹ From the analysis of the UV transient absorption measured after an ultrafast T-jump (70 °C), two rise times were observed: an ultrafast rise (<20 ps) due to the rotational and transitional motions of water, and the relatively slow one (~1 ns in water) due to the destacking dynamics of single strands. Furthermore, even after a T-jump to 70 °C, which is much higher than the melting temperature of DNA hairpin, the fluorescence intensity of the fluorescent marker

was increased in water and buffer. This result means the existence of a new intermediate melted state with a collapsed (end-to-end contact) structure in its melting pathways.

In addition, MD simulation studies further supported the existence of multiple intermediate structures of DNA hairpin. By MD simulations on the 5'-GCGCAGC sequence, Kannan and Zacharias emphasized the importance of loop and stem nucleotide interactions for hairpin folding.¹⁶² Most of the intermediates, which are rapidly converted into the native hairpin structures, include a stacking of the C2 and G3 bases and are further stabilized by hydrogen bonding to the A5 base. In some simulations that do not follow the fast pathway towards the native hairpin structure, a loop motif with G3 in the *syn* conformation accumulates, resulting in a misfolded hairpin (Fig. 10b). Those misfolded hairpin structures are the main reason for the slower folding kinetics compared to a semiflexible polymer of the same size because misfolded intermediates act as long-lived trapped states in the whole hairpin folding mechanism. Moreover, Orozco and Portella recently investigated multiple folding pathways of a small DNA hairpin, d(GCGAAGC), using MD-based studies.¹⁶³ As other previous studies suggested, it was re-confirmed that there are a number of partially folded (or misfolded) hairpin structures that have slower folding pathways than the fast one of the native hairpin. Compared to the fast folding kinetics of the DNA hairpin in the sub- μ s scale, 8 cases out of 20 times of the simulation showed longer folding times (in some cases, more than 4 μ s). Those slow folding times are due to the formation of 'stable' partially folded structures which are maintained from 0.5 to 1 μ s. In addition, an increase of the temperature encourages slow routes (Arrhenius behavior) and *anti* \leftrightarrow *syn* transitions of the guanine base, which is one of the main reasons that triggers misfolded intermediates. Accordingly, the abortive routes towards misfolded intermediates are favored rather than the number of fast routes (*anti*-Arrhenius behavior) for the native DNA hairpin structure at elevated temperature.

As suggested by Orozco and Portella, the DNA hairpin folding mechanism can be described as (1) formation of the loop structure by nucleation and closure, (2) fast base pairing of the closing base pair and (3) formation of the other Watson-Crick base pairs in the stem.¹⁶³ However, misfolded intermediates are formed by mainly *anti* \leftrightarrow *syn* transitions of the base at the end of the strand and this transition can occur in any steps of the folding mechanism. Therefore, the formation of DNA hairpin cannot be explained by a simple two state model, but should consider the ensemble of intermediates that are stable for a while, resulting in slow formation of the native hairpin structure.

7. DNA triplex

DNA triplex is one of the characteristic non-B DNA conformations and consists of the Watson-Crick base pair and Hoogsteen base pair together (Fig. 1d). The mirror repeats of the homopurine-homopyrimidine stretch in the upstream regulatory regions of several genes are known to form an intramolecular triplex structure, so-called "H-DNA".^{151,164,165} This name originates from the

characteristic triplex structure that is stabilized by hydrogen ions and preferentially formed under acidic conditions. H-DNA has been proposed to play a role in gene expression, indicating its significant biological importance. In addition, a triplex forming oligonucleotide (TFO), which forms a DNA triplex structure with high selectivity, has been used for the recognition of a specific DNA sequence, gene targeting, mutagenesis and inhibition of gene activity. Further information on the DNA triplex and its biological application can be found in the recent reviews.^{164–166}

Under the physiological condition with sufficient concentrations of salts (especially Mg^{2+}),^{167,168} TFO binds to the major groove of the host duplex and then forms the triplex structure by Hoogsteen base pairing with one strand of the host duplex, resulting in base triads.¹⁶⁹ TFO should have the identical sequence of the complementary strand of the associating strand in the DNA duplex, while the direction of the TFO relies on the type of DNA triplex. Depending on the bases (pyrimidine or purine) of the TFO interacting with the purine base of the duplex, there are two kinds of DNA triplex: parallel and anti-parallel, respectively. The most well-known triads of the parallel triplex are $T \times A \cdot T$ and $C^+ \times G \cdot C$ (\times and \cdot represent Hoogsteen and Watson–Crick base pairs, respectively), whereas the anti-parallel triplex consists of $G \times G \cdot C$ and $A \times A \cdot T$ triads. The exemplary parallel triplex and respective triads are shown in Fig. 1d.

Formation and dissociation of the DNA triplex can be described as $D + S \rightleftharpoons T$ where D, S and T represent the host duplex, single-stranded TFO and triplex, respectively (Fig. 11). Therefore, one of the optical features of DNA triplex is the existence of the two melting temperatures. In most cases, low and high T_m s can be obtained by UV melting experiments, corresponding to the dissociation of TFO from the DNA duplex and dissociation of the DNA-duplex, respectively. The latter T_m is close to that of the DNA duplex alone. From the differential scanning calorimetry (DSC) measurement, Dervan and coworkers reported that the free energy changes of triplex melting were 1.3 ± 0.1 and 17.2 ± 1.2 kcal mol⁻¹ for the first ($T \rightarrow D + S_{TFO}$) and second ($D + S_{TFO} \rightarrow 2 S_{Duplex} + S_{TFO}$) transitions, respectively.¹⁶⁹ Accordingly, the triplex is thermodynamically less stable than its host duplex. Yang *et al.* obtained K_T (triplex association constant) values of the parallel triplex formation, $(6.0 \pm 1.0) \times 10^5$ and $(7.3 \pm 0.8) \times 10^5$ M⁻¹, using FRET and fluorescence anisotropy measurements, respectively, with the 11-mer TFO (TAMRA-5'-TTTTT-CTCTCT-3') and 6-FAM-labeled 25 bp duplex.¹⁷⁰ The two K_T s were clearly in good agreement with each other as well as with the previously reported values obtained by affinity cleavage titration. In addition, it was found that as TFO becomes longer and the triplex region increases, the triplex association constants become bigger, indicating that triplex formation is more favored. On the other hand, Reither and Jeltsch analyzed antiparallel triplex formation using a similar FRET system by Yang *et al.*¹⁶⁸ From the fluorescence emission measurements of two triplexes, which are composed of TFO sequences of TAMRA-5'-GGAGGGGGAGGGG-3' (TFO1) and TAMRA-5'-GGGGAGAGGGAGG-3' (TFO2) and the respective 6-FAM-labeled 21-bp duplex sequences (DS1 and DS2), equilibrium binding constants (K_T) are determined as 2.6×10^5

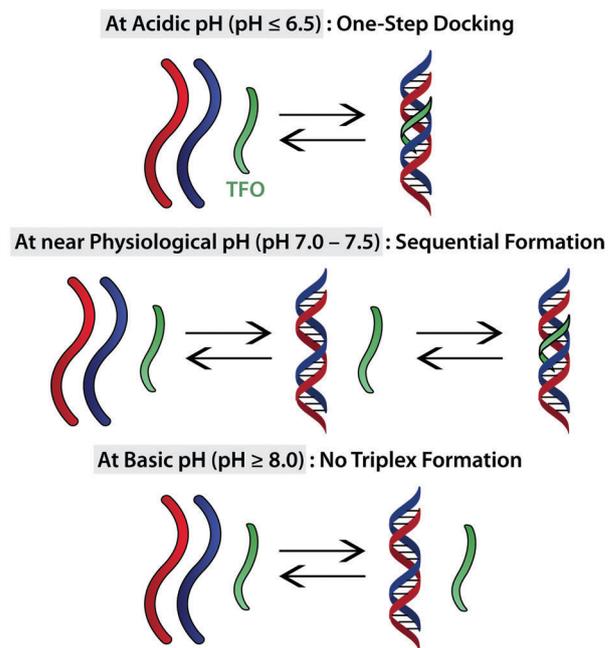


Fig. 11 Three different formation pathways of the DNA triplex as suggested by Sugimoto *et al.* (see ref. 171 for additional details).

and 2.3×10^6 M⁻¹ for TFO1-DS1 and TFO2-DS2, respectively. Considering the fact that the K_T of TFO2-DS2 is 10-fold more stable than that of TFO1-DS1, the A–AT triplex is found to be thermodynamically favorable in comparison to the G–GC triplex by approximately 5–6 kJ mol⁻¹. In addition, molecular beacon-like TFO (*i.e.*, TFO of the dye-labeled hairpin structure) was used to investigate the effect of polyamines on the formation of the DNA triplex.¹⁵³ As Mg^{2+} stabilizes the triplex structure by relieving the repulsion between backbone charges, cationic polyamines such as spermine can stabilize the DNA triplex significantly. For example, 1 μ M spermine can even induce the triplex structure under 10 mM Na⁺ condition ($K_T = 3.2 \times 10^8$ M⁻¹) where the triplex cannot be formed naturally. The synthetic analogues of spermine, such as the pentamine and hexamine, are found to be more efficient with approximately 4- and 17-fold increased K_T , respectively. In light of those studies, the formation of the DNA triplex is affected by its length, the type of triplets, and presence of cations or cationic compounds.

Moreover, the stability of the triplex is pH-dependent due to the protonation of cytosine. Sugimoto *et al.* suggested that the triplex formation at acidic pH can be described by a one-step docking model ($2 S_{Duplex} + S_{TFO} \rightleftharpoons T$) whereas the triplex at physiological pH is formed by a sequential mechanism as explained above (Fig. 11).¹⁷¹ From UV melting experiments at 20 °C and pH 6.0, the free energies of the DNA triplex, host duplex and Hoogsteen base-paired parallel duplex, composed of TFO and one strand of the host duplex, are determined as -24.58 ± 0.50 , -13.39 ± 0.16 and -8.97 ± 0.06 kcal mol⁻¹, respectively. Interestingly, $\Delta G_{\text{triplex}} < (\Delta G_{\text{host duplex}} + \Delta G_{\text{parallel duplex}})$ indicates that a larger energy would be required to disrupt the triplex compared to individually disrupt the host and parallel duplexes. In other words, the triplex at acidic pH is a more stable complex than expected,

and it undergoes the one-step dissociation. Under the same conditions (20 °C and pH 6.0), association and dissociation rate constants (k_a and k_d) of the TFO are found to be $(1.98 \pm 0.24) \times 10^3 \text{ M}^{-1} \text{ s}^{-1}$ and $(4.09 \pm 0.96) \times 10^4 \text{ s}^{-1}$, respectively, and the free energy is $-8.96 \pm 0.12 \text{ kcal mol}^{-1}$, comparable to $\Delta G_{\text{parallel duplex}}$ above. This finding implies that the formation of the Hoogsteen paired strand is independent of the structural forms of the host strands. In addition, using AFM and FRET techniques, Chang *et al.* reported that the maximum rupture force of the 17-mer DNA triplex was found at pH 4.65, suggesting the formation of a fully protonated DNA triplex.¹⁷²

Not only the aspects discussed above, but there are also various factors that affect the formation of the DNA triplex, such as the introduction of locked nucleic acids,¹⁷³ comb-type cationic copolymers,¹⁷⁴ metal ions,¹⁷⁵ molecular binders,^{176,177} molecular crowding condition,^{178,179} special TFO based on the polyamide backbone (so-called PNA)¹⁸⁰ and so on.

8. Conclusion and outlook

Although non-B DNAs such as tetraplex, Z-DNA, and A-motif were not directly detected *in vivo*, they have been proposed to participate in several biologically important processes, including the regulation, evolution, and human disease. Substantially, it is well-known that non-B DNA-forming sequences induce the genetic instability and consequently are associated with human diseases such as myotonic dystrophy (DM), Huntington's disease, fragile X syndrome (FRAX), and Friedreich's ataxia (FRDA).

In addition, some of the non-B DNAs are expected to act as a signpost of oncogene and a controller for the oncogene expression at the transcription level in the future because non-B DNA-forming sequences are frequently observed in and near the promoter region of oncogene and human telomeric DNA. Furthermore, it is well-known that B-DNA is an excellent material for building artificial nanostructures in nanotechnology, material science, molecular computing and bio-analysis because of the rigorous Watson–Crick base pairing to make the hybridization between DNA strands highly predictable, the well-defined double-helix structure and its structural stiffness and flexibility. Like B-DNA, non-B DNAs are also regarded as a fascinating material for nanotechnology because they have a unique structure, do not produce any toxic byproducts and are robust enough for the repetitive working cycle. Especially, tetraplex DNA with a higher order structure is regarded as a fascinating material for nano-machine application^{17,181–183} as well as the drug delivery system.^{184–187} Indeed, Shieh *et al.* reported that the G-quadruplex-TMPyP4 complex destroys cancer cells selectively under the 435 nm irradiation. This means that G-quadruplex can be used as a drug carrier for photodynamic therapy for cancer cells.¹⁸⁵ Moreover, there were several attempts that exploit the G-quadruplex structure for medical application.^{188,189} The applications of G-quadruplex as therapeutic targets as well as a drug carrier were excellently summarized in recent reviews.^{72,186} In addition, Modi *et al.* developed the i-motif-based nano-machine to map the spatial and temporal pH changes associated with endosome maturation.¹⁸³ These studies suggest the great potential of non-B DNA

scaffolds responsive to complex triggers in sensing, diagnostics and targeted therapies in a living cell.

In this respect, understanding their conformational changes occurring in physiological conditions is of critical importance because the conformations of various biomolecules are intimately related with their biological functions. Despite numerous studies on non-B DNA in terms of the molecular mechanism, however, it requires detailed knowledge of their conformational dynamics. For instance, the conformational dynamics of non-B DNA reported until now have shown very diverse and inconsistent results. However, most of the recent studies pointed out that the folding and unfolding dynamics of non-B DNA proceed along a multi-pathway with detectable intermediates rather than a simple two-state process. Therefore, to provide further convincing and comprehensive assignments of kinetics and conformational dynamics of non-B DNAs, a combination of single-molecule and ensemble-averaged spectroscopy is highly desirable.

Recent advances in optical imaging and biomechanical techniques offer opportunities to observe the dynamic behavior of a single biomolecule, elucidate the conformational dynamics at the level of an individual molecule and explore heterogeneity in the conformational dynamics observed between the single molecule level and in the bulk phase. In practice, the conformational dynamics and the biological functions of non-B DNAs at the single-molecule level have been extensively studied *in vitro* and *in vivo* to obtain new information and a better understanding on the conformational changes related to their biological functions that are not accessible in ensemble measurements. The single molecule study with high temporal and spatial resolution should lead to significant advances in understanding the conformational dynamics and the biological functions of non-B DNAs *in vitro* and *in vivo*.

Acknowledgements

The authors would like to thank Sooyeon Kim for her help in the data collection. T.M. thanks WCU (World Class University) program through the National Research Foundation of Korea funded by the Ministry of Education, Science and Technology (R31-10035) for the support. This work has been partly supported by a Grant-in-Aid for Scientific Research (Projects 22245022 and others) from the Ministry of Education, Culture, Sports, Science and Technology (MEXT) of Japanese Government.

References

- J. D. Watson and F. H. Crick, *Nature*, 1953, **171**, 737–738.
- E. S. Lander, L. M. Linton, B. Birren, C. Nusbaum, M. C. Zody, J. Baldwin, K. Devon, K. Dewar, M. Doyle, W. FitzHugh, R. Funke, D. Gage, K. Harris, A. Heaford, J. Howland, L. Kann, J. Lehoczy, R. LeVine, P. McEwan, K. McKernan, J. Meldrim, J. P. Mesirov, C. Miranda, W. Morris, J. Naylor, C. Raymond, M. Rosetti, R. Santos, A. Sheridan, C. Sougnez, N. Stange-Thomann, N. Stojanovic, A. Subramanian, D. Wyman, J. Rogers, J. Sulston, R. Ainscough, S. Beck, D. Bentley, J. Burton, C. Clee, N. Carter, A. Coulson, R. Deadman, P. Deloukas, A. Dunham, I. Dunham, R. Durbin, L. French, D. Grafham, S. Gregory, T. Hubbard, S. Humphray, A. Hunt, M. Jones, C. Lloyd, A. McMurray, L. Matthews, S. Mercer, S. Milne, J. C. Mullikin, A. Mungall,

- R. Plumb, M. Ross, R. Shownken, S. Sims, R. H. Waterston, R. K. Wilson, L. W. Hillier, J. D. McPherson, M. A. Marra, E. R. Mardis, L. A. Fulton, A. T. Chinwalla, K. H. Pepin, W. R. Gish, S. L. Chissoe, M. C. Wendl, K. D. Delehaunty, T. L. Miner, A. Delehaunty, J. B. Kramer, L. L. Cook, R. S. Fulton, D. L. Johnson, P. J. Minx, S. W. Clifton, T. Hawkins, E. Branscomb, P. Predki, P. Richardson, S. Wenning, T. Slezak, N. Doggett, J. F. Cheng, A. Olsen, S. Lucas, C. Elkin, E. Uberbacher, M. Frazier, R. A. Gibbs, D. M. Muzny, S. E. Scherer, J. B. Bouck, E. J. Sodergren, K. C. Worley, C. M. Rives, J. H. Gorrell, M. L. Metzker, S. L. Naylor, R. S. Kucherlapati, D. L. Nelson, G. M. Weinstock, Y. Sakaki, A. Fujiyama, M. Hattori, T. Yada, A. Toyoda, T. Itoh, C. Kawagoe, H. Watanabe, Y. Totoki, T. Taylor, J. Weissenbach, R. Heilig, W. Sauring, F. Artiguenave, P. Brottier, T. Bruls, E. Pelletier, C. Robert, P. Wincker, D. R. Smith, L. Doucette-Stamm, M. Rubinfeld, K. Weinstock, H. M. Lee, J. Dubois, A. Rosenthal, M. Platzer, G. Nyakatura, S. Taudien, A. Rump, H. Yang, J. Yu, J. Wang, G. Huang, J. Gu, L. Hood, L. Rowen, A. Madan, S. Qin, R. W. Davis, N. A. Federspiel, A. P. Abola, M. J. Proctor, R. M. Myers, J. Schmutz, M. Dickson, J. Grimwood, D. R. Cox, M. V. Olson, R. Kaul, N. Shimizu, K. Kawasaki, S. Minoshima, G. A. Evans, M. Athanasiou, R. Schultz, B. A. Roe, F. Chen, H. Pan, J. Ramser, H. Lehrach, R. Reinhardt, W. R. McCombie, M. de la Bastide, N. Dedhia, H. Blocker, K. Hornischer, G. Nordtsiek, R. Agarwala, L. Aravind, J. A. Bailey, A. Bateman, S. Batzoglou, E. Birney, P. Bork, D. G. Brown, C. B. Burge, L. Cerutti, H. C. Chen, D. Church, M. Clamp, R. R. Copley, T. Doerks, S. R. Eddy, E. E. Eichler, T. S. Furey, J. Galagan, J. G. Gilbert, C. Harmon, Y. Hayashizaki, D. Haussler, H. Hermjakob, K. Hokamp, W. Jang, L. S. Johnson, T. A. Jones, S. Kasif, A. Kasprzyk, S. Kennedy, W. J. Kent, P. Kitts, E. V. Koonin, I. Korf, D. Kulp, D. Lancet, T. M. Lowe, A. McLysaght, T. Mikkelsen, J. V. Moran, N. Mulder, V. J. Pollara, C. P. Ponting, G. Schuler, J. Schultz, G. Slater, A. F. Smit, E. Stupka, J. Szustakowski, D. Thierry-Mieg, J. Thierry-Mieg, L. Wagner, J. Wallis, R. Wheeler, A. Williams, Y. I. Wolf, K. H. Wolfe, S. P. Yang, R. F. Yeh, F. Collins, M. S. Guyer, J. Peterson, A. Felsenfeld, K. A. Wetterstrand, A. Patrino, M. J. Morgan, P. de Jong, J. J. Catanese, K. Osoegawa, H. Shizuya, S. Choi and Y. J. Chen, *Nature*, 2001, **409**, 860–921.
- 3 J. Zhao, A. Bacolla, G. Wang and K. M. Vasquez, *Cell. Mol. Life Sci.*, 2010, **67**, 43–62.
- 4 G. Wang and K. M. Vasquez, *Mutat. Res.*, 2006, **598**, 103–119.
- 5 A. Bacolla and R. D. Wells, *Mol. Carcinog.*, 2009, **48**, 273–285.
- 6 R. D. Wells, R. Dere, M. L. Hebert, M. Napierala and L. S. Son, *Nucleic Acids Res.*, 2005, **33**, 3785–3798.
- 7 Y. Wu and R. M. Brosh, Jr., *FEBS J.*, 2010, **277**, 3470–3488.
- 8 V. K. Yadav, J. K. Abraham, P. Mani, R. Kulshrestha and S. Chowdhury, *Nucleic Acids Res.*, 2008, **36**, D381–D385.
- 9 P. Jenjaroenpun and V. A. Kuznetsov, *BMC Genomics*, 2009, **10**(Suppl 3), S9.
- 10 G. Benson, *Nucleic Acids Res.*, 1999, **27**, 573–580.
- 11 R. Z. Cer, K. H. Bruce, U. S. Mudunuri, M. Yi, N. Volfovsky, B. T. Luke, A. Bacolla, J. R. Collins and R. M. Stephens, *Nucleic Acids Res.*, 2011, **39**, D383–D391.
- 12 E. Cubero, N. G. A. Abrescia, J. A. Subirana, F. J. Luque and M. Orozco, *J. Am. Chem. Soc.*, 2003, **125**, 14603–14612.
- 13 N. G. A. Abrescia, A. Thompson, T. Huynh-Dinh and J. A. Subirana, *Proc. Natl. Acad. Sci. U. S. A.*, 2002, **99**, 2806–2811.
- 14 J. A. Subirana, J. Pous, L. Urfp, C. Gouyette, J. Navaza and J. L. Campos, *J. Am. Chem. Soc.*, 2008, **130**, 6755–6760.
- 15 H. M. Al-Hashimi, E. N. Nikolova, E. Kim, A. A. Wise, P. J. O'Brien and I. Andricioaei, *Nature*, 2011, **470**, 498.
- 16 C. Lin, Y. Liu and H. Yan, *Biochemistry*, 2009, **48**, 1663–1674.
- 17 Y. Krishnan and F. C. Simmel, *Angew. Chem., Int. Ed.*, 2011, **50**, 3124–3156.
- 18 C. Teller and I. Willner, *Curr. Opin. Biotechnol.*, 2010, **21**, 376–391.
- 19 S. Modi, D. Bhatia, F. C. Simmel and Y. Krishnan, *J. Phys. Chem. Lett.*, 2010, **1**, 1994–2005.
- 20 J. T. Davis, *Angew. Chem., Int. Ed.*, 2004, **43**, 668–698.
- 21 J. L. Huppert, *FEBS J.*, 2010, **277**, 3452–3458.
- 22 A. T. Phan, V. Kuryavyy and D. J. Patel, *Curr. Opin. Struct. Biol.*, 2006, **16**, 288–298.
- 23 T. A. Brooks, S. Kendrick and L. Hurley, *FEBS J.*, 2010, **277**, 3459–3469.
- 24 T. A. Brooks and L. H. Hurley, *Nat. Rev. Cancer*, 2009, **9**, 849–861.
- 25 D. Hanahan and R. A. Weinberg, *Cell (Cambridge, Mass.)*, 2000, **100**, 57–70.
- 26 S. Schoeffner and M. A. Blasco, *Nat. Cell Biol.*, 2008, **10**, 228–236.
- 27 C. M. Azzalin, P. Reichenbach, L. Khoriauli, E. Giulotto and J. Lingner, *Science*, 2007, **318**, 798–801.
- 28 V. Viglasky, K. Tluczkova and L. Bauer, *Eur. Biophys. J.*, 2011, **40**, 29–37.
- 29 J. B. Chaires, *FEBS J.*, 2010, **277**, 1098–1106.
- 30 N. Smargiasso, F. Rosu, W. Hsia, P. Colson, E. S. Baker, M. T. Bowers, E. De Pauw and V. Gabelica, *J. Am. Chem. Soc.*, 2008, **130**, 10208–10216.
- 31 B. Heddi and A. T. Phan, *J. Am. Chem. Soc.*, 2011, **133**, 9824–9833.
- 32 J. Dai, M. Carver, C. Punchihewa, R. A. Jones and D. Yang, *Nucleic Acids Res.*, 2007, **35**, 4927–4940.
- 33 A. T. Phan, V. Kuryavyy, K. N. Luu and D. J. Patel, *Nucleic Acids Res.*, 2007, **35**, 6517–6525.
- 34 Y. Wang and D. J. Patel, *Structure*, 1993, **1**, 263–282.
- 35 F. W. Kotch, J. C. Fettingter and J. T. Davis, *Org. Lett.*, 2000, **2**, 3277–3280.
- 36 T. Li, E. Wang and S. Dong, *J. Am. Chem. Soc.*, 2009, **131**, 15082–15083.
- 37 T. Li, S. Dong and E. Wang, *J. Am. Chem. Soc.*, 2010, **132**, 13156–13157.
- 38 R. D. Gray and J. B. Chaires, *Nucleic Acids Res.*, 2008, **36**, 4191–4203.
- 39 R. D. Gray, J. Li and J. B. Chaires, *J. Phys. Chem. B*, 2009, **113**, 2676–2683.
- 40 M. M. Dailey, M. C. Miller, P. J. Bates, A. N. Lane and J. O. Trent, *Nucleic Acids Res.*, 2010, **38**, 4877–4888.
- 41 R. P. Fahlman, M. Hsing, C. S. Sporer-Tuhten and D. Sen, *Nano Lett.*, 2003, **3**, 1073–1078.
- 42 E. A. Venczel and D. Sen, *Biochemistry*, 1993, **32**, 6220–6228.
- 43 Y. C. Huang and D. Sen, *J. Am. Chem. Soc.*, 2010, **132**, 2663–2671.
- 44 C. C. Chang, C. W. Chien, Y. H. Lin, C. C. Kang and T. C. Chang, *Nucleic Acids Res.*, 2007, **35**, 2846–2860.
- 45 C. T. Lin, T. Y. Tseng, Z. F. Wang and T. C. Chang, *J. Phys. Chem. B*, 2011, **115**, 2360–2370.
- 46 J. T. Nielsen, K. Arar and M. Petersen, *Angew. Chem., Int. Ed.*, 2009, **48**, 3099–3103.
- 47 D. Pradhan, L. H. Hansen, B. Vester and M. Petersen, *Chem.–Eur. J.*, 2011, **17**, 2405–2413.
- 48 H. Saneyoshi, S. Mazzini, A. Avino, G. Portella, C. Gonzalez, M. Orozco, V. E. Marquez and R. Eritja, *Nucleic Acids Res.*, 2009, **37**, 5589–5601.
- 49 P. L. Tran, A. Virgilio, V. Esposito, G. Citarella, J. L. Mergny and A. Galeone, *Biochimie*, 2011, **93**, 399–408.
- 50 J. Lopez de la Osa, C. Gonzalez, R. Gargallo, M. Rueda, E. Cubero, M. Orozco, A. Avino and R. Eritja, *ChemBioChem*, 2006, **7**, 46–48.
- 51 J. Gros, A. Avino, J. Lopez de la Osa, C. Gonzalez, L. Lacroix, A. Perez, M. Orozco, R. Eritja and J. L. Mergny, *Chem. Commun.*, 2008, 2926–2928.
- 52 A. T. Phan and D. J. Patel, *J. Am. Chem. Soc.*, 2003, **125**, 15021–15027.
- 53 S. T. Hsu, P. Varnai, A. Bugaut, A. P. Reszka, S. Neidle and S. Balasubramanian, *J. Am. Chem. Soc.*, 2009, **131**, 13399–13409.
- 54 L. Ying, J. J. Green, H. Li, D. Klenerman and S. Balasubramanian, *Proc. Natl. Acad. Sci. U. S. A.*, 2003, **100**, 14629–14634.
- 55 P. S. Shirude, B. Okumus, L. Ying, T. Ha and S. Balasubramanian, *J. Am. Chem. Soc.*, 2007, **129**, 7484–7485.
- 56 J. Y. Lee, B. Okumus, D. S. Kim and T. Ha, *Proc. Natl. Acad. Sci. U. S. A.*, 2005, **102**, 18938–18943.
- 57 K. Okamoto, Y. Sannohe, T. Mashimo, H. Sugiyama and M. Terazima, *Bioorg. Med. Chem.*, 2008, **16**, 6873–6879.

- 58 J. W. Shim, Q. Tan and L. Q. Gu, *Nucleic Acids Res.*, 2009, **37**, 972–982.
- 59 T. Mashimo, H. Yagi, Y. Sannohe, A. Rajendran and H. Sugiyama, *J. Am. Chem. Soc.*, 2010, **132**, 14910–14918.
- 60 Y. Xu, H. Sato, Y. Sannohe, K. Shinohara and H. Sugiyama, *J. Am. Chem. Soc.*, 2008, **130**, 16470–16471.
- 61 C. Bardin and J. L. Leroy, *Nucleic Acids Res.*, 2008, **36**, 477–488.
- 62 F. Rosu, V. Gabelica, H. Poncelet and E. De Pauw, *Nucleic Acids Res.*, 2010, **38**, 5217–5225.
- 63 N. Spackova, I. Berger and J. Sponer, *J. Am. Chem. Soc.*, 1999, **121**, 5519–5534.
- 64 B. Pagano, C. A. Mattia, L. Cavallo, S. Uesugi, C. Giancola and F. Fraternali, *J. Phys. Chem. B*, 2008, **112**, 12115–12123.
- 65 D. Renciuik, I. Kejnovska, P. Skolakova, K. Bednarova, J. Motlova and M. Vorlickova, *Nucleic Acids Res.*, 2009, **37**, 6625–6634.
- 66 Z. Zhang, J. Dai, E. Veliath, R. A. Jones and D. Yang, *Nucleic Acids Res.*, 2010, **38**, 1009–1021.
- 67 M. Gueron and J. L. Leroy, *Curr. Opin. Struct. Biol.*, 2000, **10**, 326–331.
- 68 J. L. Huppert and S. Balasubramanian, *Nucleic Acids Res.*, 2007, **35**, 406–413.
- 69 S. Rankin, A. P. Reszka, J. Huppert, M. Zloh, G. N. Parkinson, A. K. Todd, S. Ladame, S. Balasubramanian and S. Neidle, *J. Am. Chem. Soc.*, 2005, **127**, 10584–10589.
- 70 A. T. Phan and J. L. Mergny, *Nucleic Acids Res.*, 2002, **30**, 4618–4625.
- 71 S. Ahmed, A. Kintanar and E. Henderson, *Nat. Struct. Biol.*, 1994, **1**, 83–88.
- 72 S. Balasubramanian and S. Neidle, *Curr. Opin. Chem. Biol.*, 2009, **13**, 345–353.
- 73 J.-L. Mergny, L. Lacroix, X. Han, J.-L. Leroy and C. Helene, *J. Am. Chem. Soc.*, 1995, **117**, 8887–8898.
- 74 S. Kendrick, Y. Akiyama, S. M. Hecht and L. H. Hurley, *J. Am. Chem. Soc.*, 2009, **131**, 17667–17676.
- 75 J. Zhou, C. Wei, G. Jia, X. Wang, Z. Feng and C. Li, *Mol. BioSyst.*, 2010, **6**, 580–586.
- 76 D. Miyoshi, H. Karimata and N. Sugimoto, *J. Am. Chem. Soc.*, 2006, **128**, 7957–7963.
- 77 S. R. Shin, K. S. Jin, C. K. Lee, S. I. Kim, G. M. Spinks, I. So, J.-H. Jeon, T. M. Kang, J. Y. Mun, S.-S. Han, M. Ree and S. J. Kim, *Adv. Mater.*, 2009, **21**, 1907–1910.
- 78 X. Li, Y. Peng, J. Ren and X. Qu, *Proc. Natl. Acad. Sci. U. S. A.*, 2006, **103**, 19658–19663.
- 79 O. Y. Fedoroff, A. Rangan, V. V. Chemeris and L. H. Hurley, *Biochemistry*, 2000, **39**, 15083–15090.
- 80 X. Han, J. L. Leroy and M. Gueron, *J. Mol. Biol.*, 1998, **278**, 949–965.
- 81 S. Nonin, A. T. Phan and J. L. Leroy, *Structure*, 1997, **5**, 1231–1246.
- 82 S. Nonin and J. L. Leroy, *J. Biomol. Struct. Dyn.*, 1999, **16**, 1354.
- 83 K. S. Jin, S. R. Shin, B. Ahn, Y. Rho, S. J. Kim and M. Ree, *J. Phys. Chem. B*, 2009, **113**, 1852–1856.
- 84 J. M. Dettler, R. Buscaglia, J. Cui, D. Cashman, M. Blynn and E. A. Lewis, *Biophys. J.*, 2010, **99**, 561–567.
- 85 A. Rajendran, S. Nakano and N. Sugimoto, *Chem. Commun.*, 2010, **46**, 1299–1301.
- 86 S. Dhakal, J. D. Schonhoft, D. Koirala, Z. Yu, S. Basu and H. Mao, *J. Am. Chem. Soc.*, 2010, **132**, 8991–8997.
- 87 T. E. Malliavin, J. Gau, K. Snoussi and J. L. Leroy, *Biophys. J.*, 2003, **84**, 3838–3847.
- 88 T. E. Malliavin, K. Snoussi and J. L. Leroy, *Magn. Reson. Chem.*, 2003, **41**, 18–25.
- 89 J. L. Leroy, K. Snoussi and M. Gueron, *Magn. Reson. Chem.*, 2001, **39**, S171–S176.
- 90 J. L. Leroy, M. Gueron, J. L. Mergny and C. Helene, *Nucleic Acids Res.*, 1994, **22**, 1600–1606.
- 91 Y. Zhao, Z. X. Zeng, Z. Y. Kan, Y. H. Hao and Z. Tan, *ChemBioChem*, 2005, **6**, 1957–1960.
- 92 D. S. Liu and S. Balasubramanian, *Angew. Chem., Int. Ed.*, 2003, **42**, 5734–5736.
- 93 Y. Mao, D. Liu, S. Wang, S. Luo, W. Wang, Y. Yang, Q. Ouyang and L. Jiang, *Nucleic Acids Res.*, 2007, **35**, e33.
- 94 A. H. Wang, G. J. Quigley, F. J. Kolpak, J. L. Crawford, J. H. van Boom, G. van der Marel and A. Rich, *Nature*, 1979, **282**, 680–686.
- 95 B. H. Johnston, *Methods Enzymol.*, 1992, **211**, 127–158.
- 96 T. Kimura, K. Kawai and T. Majima, *Chem. Commun.*, 2004, 268–269.
- 97 T. Kimura, K. Kawai and T. Majima, *Chem. Commun.*, 2006, 1542–1544.
- 98 P. Khuu, M. Sandor, J. DeYoung and P. S. Ho, *Proc. Natl. Acad. Sci. U. S. A.*, 2007, **104**, 16528–16533.
- 99 B. K. Ray, S. Dhar, A. Shakya and A. Ray, *Proc. Natl. Acad. Sci. U. S. A.*, 2011, **108**, 103–108.
- 100 B. Wittig, S. Wolf, T. Dorbic, W. Vahrson and A. Rich, *EMBO J.*, 1992, **11**, 4653–4663.
- 101 D. B. Oh, Y. G. Kim and A. Rich, *Proc. Natl. Acad. Sci. U. S. A.*, 2002, **99**, 16666–16671.
- 102 S. Zhang, C. Lockshin, A. Herbert, E. Winter and A. Rich, *EMBO J.*, 1992, **11**, 3787–3796.
- 103 A. Herbert, J. Alfken, Y. G. Kim, I. S. Mian, K. Nishikura and A. Rich, *Proc. Natl. Acad. Sci. U. S. A.*, 1997, **94**, 8421–8426.
- 104 T. Schwartz, M. A. Rould, K. Lowenhaupt, A. Herbert and A. Rich, *Science*, 1999, **284**, 1841–1845.
- 105 J. A. Kwon and A. Rich, *Proc. Natl. Acad. Sci. U. S. A.*, 2005, **102**, 12759–12764.
- 106 S. Feng, H. Li, J. Zhao, K. Pervushin, K. Lowenhaupt, T. U. Schwartz and P. Droge, *Nat. Struct. Mol. Biol.*, 2011, **18**, 169–176.
- 107 S. C. Ha, K. Lowenhaupt, A. Rich, Y. G. Kim and K. K. Kim, *Nature*, 2005, **437**, 1183–1186.
- 108 J. R. Bothe, K. Lowenhaupt and H. M. Al-Hashimi, *J. Am. Chem. Soc.*, 2011, **133**, 2016–2018.
- 109 F. M. Pohl and T. M. Jovin, *J. Mol. Biol.*, 1972, **67**, 375–396.
- 110 A. Rich and S. Zhang, *Nat. Rev. Genet.*, 2003, **4**, 566–572.
- 111 A. Herbert and A. Rich, *J. Biol. Chem.*, 1996, **271**, 11595–11598.
- 112 N. Shimada, A. Kano and A. Maruyama, *Adv. Funct. Mater.*, 2009, **19**, 3590–3595.
- 113 R. Tashiro and H. Sugiyama, *Angew. Chem., Int. Ed.*, 2003, **42**, 6018–6020.
- 114 R. Tashiro and H. Sugiyama, *J. Am. Chem. Soc.*, 2005, **127**, 2094–2097.
- 115 L. J. Peck, A. Nordheim, A. Rich and J. C. Wang, *Proc. Natl. Acad. Sci. U. S. A.*, 1982, **79**, 4560–4564.
- 116 D. B. Haniford and D. E. Pulleyblank, *J. Biomol. Struct. Dyn.*, 1983, **1**, 593–609.
- 117 D. A. Koster, A. Crut, S. Shuman, M. A. Bjornsti and N. H. Dekker, *Cell (Cambridge, Mass.)*, 2010, **142**, 519–530.
- 118 A. I. Murchie and D. M. Lilley, *Methods Enzymol.*, 1992, **211**, 158–180.
- 119 M. Lee, S. H. Kim and S. C. Hong, *Proc. Natl. Acad. Sci. U. S. A.*, 2010, **107**, 4985–4990.
- 120 T. R. Strick, J. F. Allemand, D. Bensimon, A. Bensimon and V. Croquette, *Science*, 1996, **271**, 1835–1837.
- 121 A. Nordheim, E. M. Lafer, L. J. Peck, J. C. Wang, B. D. Stollar and A. Rich, *Cell (Cambridge, Mass.)*, 1982, **31**, 309–318.
- 122 L. J. Peck and J. C. Wang, *Proc. Natl. Acad. Sci. U. S. A.*, 1983, **80**, 6206–6210.
- 123 I. Mamajanov, A. E. Engelhart, H. D. Bean and N. V. Hud, *Angew. Chem., Int. Ed.*, 2010, **49**, 6310–6314.
- 124 A. G. Herbert and A. Rich, *Nucleic Acids Res.*, 1993, **21**, 2669–2672.
- 125 J. Geng, C. Zhao, J. Ren and X. Qu, *Chem. Commun.*, 2010, **46**, 7187–7189.
- 126 J. L. Breslow, J. McPherson, A. L. Nussbaum, H. W. Williams, F. Lofquist-Kahl, S. K. Karathanasis and V. I. Zannis, *J. Biol. Chem.*, 1982, **257**, 14639–14641.
- 127 E. I. Rogaev, R. Sherrington, C. Wu, G. Levesque, Y. Liang, E. A. Rogaeva, M. Ikeda, K. Holman, C. Lin, W. J. Lukiw, P. J. de Jong, P. E. Fraser, J. M. Rommens and P. St George-Hyslop, *Genomics*, 1997, **40**, 145–147.
- 128 S. Bae, D. Kim, K. K. Kim, Y. G. Kim and S. Hohng, *J. Am. Chem. Soc.*, 2011, **133**, 668–671.
- 129 H. Sugiyama, K. Kawai, A. Matsunaga, K. Fujimoto, I. Saito, H. Robinson and A. H. Wang, *Nucleic Acids Res.*, 1996, **24**, 1272–1278.
- 130 S. Chakraborty, S. Sharma, P. K. Maiti and Y. Krishnan, *Nucleic Acids Res.*, 2009, **37**, 2810–2817.
- 131 P. Giri and G. Suresh Kumar, *Mol. BioSyst.*, 2010, **6**, 81–88.
- 132 O. P. Cetinkol and N. V. Hud, *Nucleic Acids Res.*, 2009, **37**, 611–621.

- 133 H. Xi, D. Gray, S. Kumar and D. P. Arya, *FEBS Lett.*, 2009, **583**, 2269–2275.
- 134 P. Giri and G. S. Kumar, *Arch. Biochem. Biophys.*, 2008, **474**, 183–192.
- 135 B. Alberts, *Molecular biology of the cell*, Garland Science, New York, 4th edn, 2002.
- 136 A. Rich, D. R. Davies, F. H. Crick and J. D. Watson, *J. Mol. Biol.*, 1961, **3**, 71–86.
- 137 B. Janik, R. G. Sommer and A. M. Bobst, *Biochim. Biophys. Acta*, 1972, **281**, 152–168.
- 138 R. Maggini, F. Secco, M. Venturini and H. Diebler, *J. Chem. Soc., Faraday Trans.*, 1994, **90**, 2359–2363.
- 139 M. I. Zarudnaya, *Mol. Biol.*, 1998, **32**, 417–422.
- 140 A. G. Petrovic and P. L. Polavarapu, *J. Phys. Chem. B*, 2005, **109**, 23698–23705.
- 141 C. E. Crespo-Hernández and B. Kohler, *J. Phys. Chem. B*, 2004, **108**, 11182–11188.
- 142 K. Bhadra and G. S. Kumar, *Biochim. Biophys. Acta*, 2011, **1810**, 485–496.
- 143 F. Xing, G. Song, J. Ren, J. B. Chaires and X. Qu, *FEBS Lett.*, 2005, **579**, 5035–5039.
- 144 M. Maiti and G. S. Kumar, *J. Nucleic Acids*, 2010, **2010**, 593408.
- 145 O. Persil, C. T. Santai, S. S. Jain and N. V. Hud, *J. Am. Chem. Soc.*, 2004, **126**, 8644–8645.
- 146 I. S. Joung, Ö. Persil Çetinkol, N. V. Hud and T. E. Cheatham, *Nucleic Acids Res.*, 2009, **37**, 7715–7727.
- 147 A. N. Gough, R. L. Jones and W. D. Wilson, *J. Med. Chem.*, 1979, **22**, 1551–1554.
- 148 P. Giri and G. S. Kumar, *J. Photochem. Photobiol., A*, 2008, **194**, 111–121.
- 149 S. S. Jain, M. Polak and N. V. Hud, *Nucleic Acids Res.*, 2003, **31**, 4608–4615.
- 150 M. T. Madigan, J. M. Martinko and T. D. Brock, *Brock biology of microorganisms*, Pearson Prentice Hall, Upper Saddle River, NJ, 11th edn, 2006.
- 151 M. D. Frank-Kamenetskii, *Methods Enzymol.*, 1992, **211**, 180–191.
- 152 S. Cao and S.-J. Chen, *Nucleic Acids Res.*, 2006, **34**, 2634–2652.
- 153 T. Antony, T. Thomas, L. H. Sigal, A. Shirahata and T. J. Thomas, *Biochemistry*, 2001, **40**, 9387–9395.
- 154 S. Tyagi and F. R. Kramer, *Nat. Biotechnol.*, 1996, **14**, 303–308.
- 155 G. Bonnet, S. Tyagi, A. Libchaber and F. R. Kramer, *Proc. Natl. Acad. Sci. U. S. A.*, 1999, **96**, 6171–6176.
- 156 J. Liu, Z. Cao and Y. Lu, *Chem. Rev.*, 2009, **109**, 1948–1998.
- 157 S. Yoshizawa, G. Kawai, K. Watanabe, K. Miura and I. Hirao, *Biochemistry*, 1997, **36**, 4761–4767.
- 158 S. Kannan and M. Zacharias, *Nucleic Acids Res.*, 2011, DOI: 10.1093/nar/gkr541.
- 159 G. Bonnet, O. Krichevsky and A. Libchaber, *Proc. Natl. Acad. Sci. U. S. A.*, 1998, **95**, 8602–8606.
- 160 M. I. Wallace, L. M. Ying, S. Balasubramanian and D. Klenerman, *J. Phys. Chem. B*, 2000, **104**, 11551–11555.
- 161 H. Ma, C. Wan, A. Wu and A. H. Zewail, *Proc. Natl. Acad. Sci. U. S. A.*, 2007, **104**, 712–716.
- 162 S. Kannan and M. Zacharias, *Biophys. J.*, 2007, **93**, 3218–3228.
- 163 M. Orozco and G. Portella, *Angew. Chem., Int. Ed.*, 2010, **49**, 7673–7676.
- 164 K. M. Vasquez and P. M. Glazer, *Q. Rev. Biophys.*, 2002, **35**, 89–107.
- 165 A. Mukherjee and K. M. Vasquez, *Biochimie*, 2011, **93**, 1197–1208.
- 166 P. P. Chan and P. M. Glazer, *J. Mol. Med. (Berlin)*, 1997, **75**, 267–282.
- 167 J. E. Lee, T. Kim, S. Y. Kim and S. W. Kim, *Chem. Phys. Lett.*, 2010, **490**, 230–233.
- 168 S. Reither and A. Jeltsch, *BMC Biochem.*, 2002, **3**, 27.
- 169 G. E. Plum, Y. W. Park, S. F. Singleton, P. B. Dervan and K. J. Breslauer, *Proc. Natl. Acad. Sci. U. S. A.*, 1990, **87**, 9436–9440.
- 170 M. S. Yang, S. S. Ghosh and D. P. Millar, *Biochemistry*, 1994, **33**, 15329–15337.
- 171 N. Sugimoto, P. Wu, H. Hara and Y. Kawamoto, *Biochemistry*, 2001, **40**, 9396–9405.
- 172 C. C. Chang, P. Y. Lin, Y. F. Chen, C. S. Chang and L. S. Kan, *Appl. Phys. Lett.*, 2007, **91**, 203901.
- 173 N. Kumar, K. E. Nielsen, S. Maiti and M. Petersen, *J. Am. Chem. Soc.*, 2005, **128**, 14–15.
- 174 A. Maruyama, M. Katoh, T. Ishihara and T. Akaike, *Bioconjugate Chem.*, 1997, **8**, 3–6.
- 175 K. Tanaka, Y. Yamada and M. Shionoya, *J. Am. Chem. Soc.*, 2002, **124**, 8802–8803.
- 176 A. K. Jain and S. Bhattacharya, *Bioconjugate Chem.*, 2010, **21**, 1389–1403.
- 177 D. P. Arya, L. Micovic, I. Charles, R. L. Coffee, B. Willis and L. Xue, *J. Am. Chem. Soc.*, 2003, **125**, 3733–3744.
- 178 D. Miyoshi and N. Sugimoto, *Biochimie*, 2008, **90**, 1040–1051.
- 179 D. Miyoshi, K. Nakamura, H. Tateishi-Karimata, T. Ohmichi and N. Sugimoto, *J. Am. Chem. Soc.*, 2009, **131**, 3522–3531.
- 180 P. Nielsen, M. Egholm, R. Berg and O. Buchardt, *Science*, 1991, **254**, 1497–1500.
- 181 C. Wang, Z. Huang, Y. Lin, J. Ren and X. Qu, *Adv. Mater.*, 2010, **22**, 2792–2798.
- 182 Y. Yang, G. Liu, H. Liu, D. Li, C. Fan and D. Liu, *Nano Lett.*, 2010, **10**, 1393–1397.
- 183 S. Modi, M. G. Swetha, D. Goswami, G. D. Gupta, S. Mayor and Y. Krishnan, *Nat. Nanotechnol.*, 2009, **4**, 325–330.
- 184 C. Chen, F. Pu, Z. Huang, Z. Liu, J. Ren and X. Qu, *Nucleic Acids Res.*, 2011, **39**, 1638–1644.
- 185 Y. A. Shieh, S. J. Yang, M. F. Wei and M. J. Shieh, *ACS Nano*, 2010, **4**, 1433–1442.
- 186 S. Neidle, *Curr. Opin. Struct. Biol.*, 2009, **19**, 239–250.
- 187 C. Xu, C. Zhao, J. Ren and X. Qu, *Chem. Commun.*, 2011, **47**, 8043–8045.
- 188 M. Hagihara, L. Yamauchi, A. Seo, K. Yoneda, M. Senda and K. Nakatani, *J. Am. Chem. Soc.*, 2010, **132**, 11171–11178.
- 189 Y. Xu, K. Ito, Y. Suzuki and M. Komiyama, *J. Am. Chem. Soc.*, 2010, **132**, 631–637.