6416–6434 *Nucleic Acids Research, 2018, Vol. 46, No. 13* *Published online 26 June 2018*
doi: 10.1093/nar/gky561

SURVEY AND SUMMARY

A survey of recent unusual high-resolution DNA structures provoked by mismatches, repeats and ligand binding

Roshan Satange[1,2,†], Chung-ke Chang[3,†] and Ming-Hon Hou[1,2,*]

[1]Institute of Genomics and Bioinformatics, National Chung Hsing University, Taichung, Taiwan, [2]Ph.D. Program in Medical Biotechnology, National Chung Hsing University, Taichung, Taiwan and [3]Institute of Biomedical Sciences, Academia Sinica, Taipei, Taiwan

Received May 07, 2018; Revised June 05, 2018; Editorial Decision June 06, 2018; Accepted June 08, 2018

ABSTRACT

The structure of the DNA duplex is arguably one of the most important biological structures elucidated in modern history. DNA duplex structure is closely associated with essential biological functions such as DNA replication and RNA transcription. In addition to the classical A-, B- and Z-DNA conformations, DNA duplexes are capable of assuming a variety of alternative conformations depending on the sequence and environmental context. A considerable number of these unusual DNA duplex structures have been identified in the past decade, and some of them have been found to be closely associated with different biological functions and pathological conditions. In this manuscript, we review a selection of unusual DNA duplex structures, particularly those originating from base pair mismatch, repetitive sequence motifs and ligand-induced structures. Although the biological significance of these novel structures has not yet been established in most cases, the illustrated conformational versatility of DNA could have relevance for pharmaceutical or nanotechnology development. A perspective on the future directions of this field is also presented.

INTRODUCTION

The DNA encodes the genetic instructions for all life processes and is required for inheritance. As a substance, DNA was isolated as early as 1878 by Albrecht Kossel, but its role in heredity was not firmly established until the 1950s. The mechanistic basis of DNA as the main agent of heredity was elucidated through a series of elegant structural studies conducted by Watson, Crick, Wilkins, Franklin and others, which established the seminal 'double-helix' structure of the DNA duplex (1–3).

The canonical DNA duplex structure proposed by Watson and Crick in 1953, now called the 'B-form' DNA or B-DNA, played a key role in explaining the replication process (4). Further studies based on this structural model eventually led to the proposition of the 'central dogma' and birth of the field of molecular biology. However, it was soon recognized that DNA was unlikely to exist in this perfect, linear B-DNA conformation within the cell. The discovery of alternative conformations of the DNA duplex, such as A-DNA and Z-DNA, and confirmation of their presence within cells opened up new venues of research which eventually linked these alternative conformations with key cellular processes (5–7). Moreover, along with A-, B- and Z-form of DNA, other DNA structures like bulges, hairpins, cruciforms, parallel-stranded DNAs, triplexes, quadruplexes, and i-motifs have also been widely studied. These alternative DNA structures might not only be important for interactions with proteins involved in replication, gene expression and recombination, but these would also have an impact on DNA damage, repair and genetic instability (8,9). Indeed, there is a growing literature of 'unusual' DNA duplex structures highlighting the dependence of DNA structural polymorphism on sequence composition, hydration status, presence of metal ions, interaction with ligands both large and small, etc.

The importance of these unusual DNA duplex structures cannot be overstated. Many of these structures can be recognized by specific proteins and may be involved in gene regulation and DNA damage repair (10). Repetitive DNA sequences are known to form unusual structures which

*To whom correspondence should be addressed. Tel: +886 4 2284 0338 (Ext. 7011); Fax: +886 4 2285 9329; Email: mhho@nchu.edu.tw
†The authors wish it to be known that, in their opinion, the first two authors should be regarded as Joint First Authors.

have been implicated in human disease progression (11,12). Viruses and plasmids also contain DNA sequences that may form unusual structures (13–15). These unusual structures are potential targets for the development of novel anticancer, antiviral or antibacterial pharmaceutics (16–20). Furthermore, binding of small molecules may also induce formation of unusual DNA structures, which may offer new venues to probe and shape DNA materials (20–22). Given their importance, it was perplexing to find that the structures of unusual DNA duplexes have rarely been reviewed for the past decade. Therefore, we believe that it was time to revisit this topic and update our current understanding of the principles behind the formation of unusual DNA duplex structures and their potential applications in light of recent advances in the field.

In the current review, we will start with a concise survey of the various structural features of unusual DNA duplexes associated with mismatches and tandem repeats. We will also describe structural details of unusual ligand-bound DNA conformations, which may provide insights into the mechanisms underlying drug action at the cellular and molecular levels. Although the current article is limited to a review of high resolution structures determined *in vitro* and deposited in the PDB, it should provide useful information about current progress in this interdisciplinary field of nucleic acid research. The aforementioned structures are summarized in Tables 1–3. In the last part of this review, we will provide a perspective on the future direction of the field and its potential applications. We hope that the collated structural information will provide the reader with a better understanding of how unusual DNA structures might affect the biology of the DNA and encourage newcomers to contribute towards this exciting and unique area of nucleic acid research.

MISMATCH-INDUCED UNUSUAL DUPLEX STRUCTURES

Structural polymorphism of DNA mismatches

Base pair mismatches may arise from a variety of mechanisms, including spontaneous cytosine methylation or deamination (G:T mismatch), transient misalignment during replication (various mismatches), and oxidative damage (8-oxoguanine mismatches) (23). Regardless of the origin, these mismatches destabilize the duplex structure to various extents depending on the type and nature of the mismatched bases. Each mismatch may cause local distortions in the DNA duplex structure, and accumulation of mismatches may eventually alter the structural topology of the DNA. It is thus not surprising that all cells have dedicated mismatch recognition and repair systems. Failure to correct for mismatched base pairs eventually lead to abnormal functioning of the cell and is the major mechanism behind an increasing number of genetic defects and cancers.

Different hypotheses for mismatch detection have been proposed, including recognition of the mispaired base through changes in the hydrogen bond donor–acceptor pattern and detection of DNA deformation such as kinking and bending. These hypotheses revolve around a thorough understanding of the structures of various mismatched DNA duplexes, many of which have been elucidated in the last century. The reader is referred to the excellent reviews

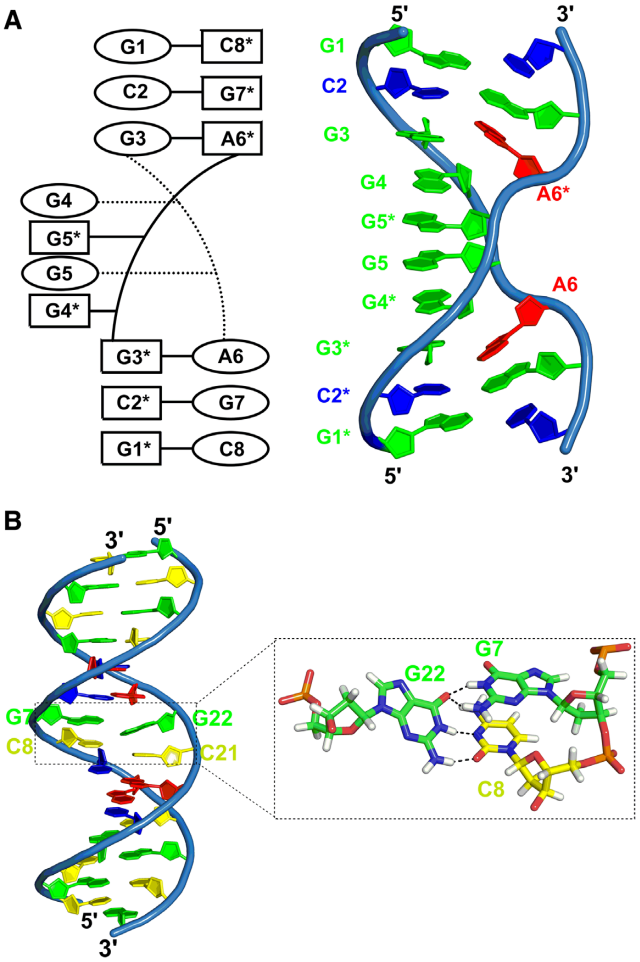

Figure 1. (**A**) Schematic diagram and overall structure of the base-intercalated duplex d(GCGGGAGC)$_2$ (PDB ID: 2FZA). (**B**) Overall structure of the DNA duplex reported by Ghosh et al (PDB ID: 2MJX). The hydrogen bond triad formed between dG7, dC8 and dG22 are shown in the inset.

by Fukui *et al.* (24–26). In general, the structures of reported mismatch-containing DNA vary greatly. The identified mismatch containing DNA structures further corroborate this fact (Table 1). For example, the structure of d(GCGGGAGC)$_2$ contains sheared G:A mismatches on both ends of the duplex which forces G4 and G5 of both strands to extrude out of the helix and intercalate with each other (27). Surprisingly, the central G:G mismatch appears to have little effect on the structure because d(GCGATAGC)$_2$, which contains only G:A mismatches, also formed base-intercalated structures involving A4 and T5 (28) (Figure 1A). The exposed bases may assist in the formation of multiplex structures and probably play a functional role inside the cell. On the other side of the spectrum is the G:G–C:C dual mismatch structure reported by Ghosh et al. (29) (Figure 1B). Its major feature is the formation of a hydrogen bond triad at the mismatch site. However, the helix structure is almost identical to that of B-DNA. The changes are subtle: in addition to local perturbations, thermodynamic studies revealed that the dual mismatch struc-

Table 1. List of unusual DNA structures and features involving base pair mismatch and repetitive DNA sequences

DNA sequence	Structural feature	PDB ID	Ref
Mismatch base pairs-containing DNA duplexes			
d(GCGGGAGC)	Sheared G:A mismatch, Base-intercalated duplex	2FZA	(27)
d(GCGCATGCTACGCG)	Double G:G and C:C mismatch	2MJX	(29)
d(CGCGATTTCGCG)	T–Hg^{II}–T mismatch	4L24	(38)
d(CGCGATTTCGCG)	T:T mismatch	4L25	(38)
d(GCCCGTGC)	C–Hg^{II}–T mismatch	5WSQ	(39)
d(GGTCGTCC)	T–Hg^{II}–T mismatch	5WSR	(39)
d(GCACGCGC)	G–Ag^{I}–G and C–Ag^{I}–C mismatch	5XJZ	(44)
d(GGACTCGACTCC)	C–Ag^{I}–C, G–Ag^{I}–G, G–Ag^{I}–C and T–Ag^{I}–T	5IX7	(45)
Repeat DNA motif structures			
d(GCATGCATGC)	Double folded structure	4ZKK	(59)
d(GTGGAATGGAAC)	G:A mismatch, Guanine zipper core	5GUN	(64)
d(CCTGCCTG)	CCTG repeats, Mini-dumbbell	5GWL	(66)
d(TTTATTTA)	TTTA repeats, Mini-dumbbell	5GWQ	(66)

Table 2. List of DNA duplex-small molecule compound complexes which cause DNA bending

Drug	DNA sequence	DNA bending type#	PDB ID	Drug binding mode	Ref
Aurelic acid-type compounds					
[Mg^{II}-(Chromomycin A_3)$_2$]	d(TTGGCCAA)	DNA bending	1VAQ	Minor grove binding	(77)
Mithramycin analogue (Mith SA-phe)	d(AGGGTACCCT)	DNA bending	5JW0	Minor grove binding	(78)
Mithramycin analogue (Mith SA-phe)	d(AGGGATCCCT)	DNA bending	5JW2	Minor grove binding	(78)
Bis-intercalator compounds					
Echinomycin	d(ACGTCGT)$_2$	Smooth bending	5YTZ	Intercalation	(83)
Echinomycin	d(ACGACGT/ACGTCGT)	Smooth bending	5YTY	Intercalation	(83)
Ruthenium-based compounds					
Lambda-[Ru(tap)$_2$(dppz)]$^{2+}$	d(TCGGCGCCGA)	DNA kink	3QRN	Intercalation and semi-intercalation	(88)
Lambda-[Ru(tap)$_2$(dppz)]$^{2+}$	d(TCGGCGCCGA)	DNA bending, DNA kink	4LTJ, 4LTF	Intercalation	(89)
Delta-[Ru(phen)$_2$(dppz]$^{2+}$	d(TCGGCGCCGA)	DNA kink	5JEV, 5JEU	Intercalation and semi-intercalation	(87)
Polyamide-based compounds					
Cyclic Py/Im polyamide	d(CCAGGCCTGG)	Smooth bending	3I5L	Minor grove binding	(94)
β-amino turn-linked cyclic Py-Im polyamide	d(CCAGTACTGG)	Smooth bending	3OMJ	Minor grove binding	(95)
Thiazole-containing hairpin polyamides	d(CGATGTACATCG)	Smooth bending	5ODF, 5ODM, 5OE1, 5OCZ	Groove binding	(101)

P.S. # - The DNA bending types are categorized based on the extent of bending of DNA axis upon ligand binding. Severe bending includes structures causing bending angles of ∼20° to ∼45°, DNA kink is categorized by bending angles >45° whereas smooth bending includes structures causing <20° bending of the DNA axis.

Table 3. List of DNA duplex-small molecule compound complexes which cause base flip-out

Drug	DNA sequence	PDB ID	Drug binding mode	Flipped-out base pair	Ref
DNA mismatch structures with flipped-out bases					
Naphthyridine-azaquinolone	d(CTAACAGAATG)	1X26	Intercalation	G:C	(102)
[Ni^{II}-(Chromomycin A_3)$_2$]	d(TT(CCG)$_3$AA)	5XEW	Minor grove binding	C:C	(103)
Δ-[Rh(bpy)$_2$chrysi]$^{3+}$	d(CGGAAATTCCCG)	2O1I	Intercalation	**A:C**	(104)
Δ-Rh(bpy)$_2$chrysi^{3+}	d(CGGACTCCG)	-	-	**C:C**	(105)
Δ-Rh(bpy)$_2$(chrysi)$^{3+}$	d(CGGAAATTACCG)	3GSK, 3GSJ	Intercalation	**A:A**	(106)
Δ-[Ru(bpy)$_2$dppz]$^{2+}$	d(CGGAAATTACCG)	4E1U	Intercalation	**A:A**	(107)
Actinomycin D	d(ATGCGGCAT)	4HIV	Intercalation	G:G	(108)
(–)-lomaiviticin A	d(GCTATAGC)	2N96	Intercalation	**A:T**	(109)
bis-naphthalene macrocycle (2,7-BisNP)	d(CGTCGTAGTGC)	2LLJ, 2LL9	Intercalation	**T:T**	(110)

P.S.- Flipped-out bases are highlighted in bold.

ture is less stable than that of *bona fide* B-DNA with perfect Watson–Crick base pairs.

Interestingly, mismatch repair proteins such as MutS often are capable of recognizing multiple types of mismatches with high fidelity even when these mismatches have different structures (30). Recent research using molecular dynamics and solution experiments suggests that the conformational dynamics, i.e. motion time scale, of the DNA containing the mismatch may be a key factor for mismatch detection efficiency (31). Indeed, mismatched bases such as T:T have been found to 'wobble' between different pairing patterns, with the preferential pairing being dependent on the context of the flanking sequences (32,33). Most DNA mismatches are destabilized compared to the B-DNA duplex, which may facilitate exchange of the local conformations. It is possible that mismatch repair proteins are capable of detecting these structural fluctuations as a means of identifying mismatch sites on the DNA.

Heavy metal ions modulate mismatch DNA structure

A recent development is the structural elucidation of various heavy metal-bound DNA containing mismatched base pairs (Table 1). Although not directly relevant to biology, metal-containing DNA has great application potential in areas such as nanotechnology and may become important tools for biological research (34–37). Of particular importance is the interaction between T:T or C:T base pairs and Hg^{II}. T:T and C:T mismatches are known to have high binding affinity towards Hg^{II} and may be useful in the detection or decontamination of mercury pollutants in the environment. However, the structural basis behind this affinity remained obscure for a long time. Kondo *et al.* reported the crystal structure of B-DNA duplex containing two consecutive Hg^{II}-mediated T:T base pairs (T–Hg^{II}–T) (38). In the absence of Hg^{II}, the DNA duplex adopts an unusual nonhelical conformation with the thymines flipped out and a hydrogen bond triad in the core region (Figure 2A). However, in the presence of Hg^{II}, the same DNA duplex adopts the B-DNA conformation (Figure 2B). The Hg^{II} ion interacts with the N3 atoms of the two thymines comprising the T:T mismatch (Figure 2C). These findings have been corroborated by subsequent NMR studies of a similar DNA sequence. The crystal structures of a series of DNA duplexes containing either Hg^{II}-mediated T:T or C:T mismatches (Figure 2D) have also been reported by Liu *et al.* (39). In contrast to the previous report, these DNA duplexes all adopted the A-DNA conformation; however, the C–Hg^{II}–T duplexes resulted in the compression of the helix axis compared to normal A-DNA containing GC-rich sequences. In T–Hg^{II}–T, the mercury ion was found to interact with the T:T mismatch in multiple modalities: in addition to the aforementioned N3–Hg^{II}–N3 interaction, Hg^{II} also may form O2–Hg^{II}–N3 and N3–Hg^{II}–O4 interactions, which highlights the dynamic nature of T–Hg^{II}–T pairings (Figure 2E and F).

Silver (Ag^{I}) is another heavy metal ion which is known to mediate base-base interactions (40,41). Whilst being more flexible in terms of which bases may interact with it, previous structural studies were conducted in the context of the helical duplex (42,43). Recently, a nonhelical conformation of DNA duplex containing a C–Ag^{I}–C pair and a G–Ag^{I}–G pair has been reported (44) (Figure 3A–D). The structure contains two sharp-turned DNA strands along with two Ag^{I} ions mediating the p(42)airing between the two guanine residues and between the two cytocine residues at the 3′ terminus, respectively. The two strands further zip together, adopting an unusual nonhelical conformation that mimics a pyramid shape in one orientation and cylinder shape in the other. This is the first known nonhelical DNA structure driven by heavy metal ions, suggesting that heavy metal ions may contribute towards the structural diversity of DNA. Kondo et al. reported the crystal structure of a metallo-DNA nanowire at a resolution of 1.39 Å composed of dodecamers held together by four different Ag-mediated base pairs (45) (Figure 3E). The two DNA dodecamers form an antiparallel right-handed helical duplex that is very similar to the standard B-form conformation of DNA. Strikingly, this DNA duplex does not contain any A:T and G:C base pairs. Instead, it has only metallobase pairs: C–Ag–C, G–Ag–G, G–Ag–C and T–Ag–T while the adenine (A) was flipped out of the duplex. Such a structure may provide novel information for the development of new materials, such as DNA nanowires, DNA-based metal ion sensors, and metal-containing DNA nanomaterials (46–48).

REPETITIVE DNA SEQUENCES

DNA sequences containing repeating elements, particularly trinucleotide repeats, are closely associated with human diseases (49–51). Expansion of trinucleotide repeats, for example, is a major cause of age-onset neurological disorders (51,52). Most repetitive DNA sequences are quasi-palindromes and may form intra-strand hairpin structures harboring A:A, T:T or C:C mismatches in addition to the normal double-stranded helix structure. Some trinucleotide repeats may even form triplex and quadruplex structures depending on the sequence (53,54). The potential to assume alternative conformations is a major cause of the genetic instability (52,55,56). It is thus not surprising that repetitive DNA sequences have received special attention from structural biologists for almost half a century. The interested reader is encouraged to seek out the numerous reviews that have been written on the subject (20,57,58). We will devote this section to highlight a few recent developments (Table 1).

Thirugnanasambandam *et al.* recently reported a novel double-folded structure of the decadeoxyribonucleotide d(GCATGCATGC) (59) (Figure 4A). The sequence is a self-complementary purine–pyrimidine repeat. The structure appears to be determined by the GCAT tetrad, which forms a local quadruplex with its symmetry-related neighbor. This region constitutes a 'bi-loop' motif previously observed in a number of sequences (60–63). The backbone reverses direction twice, making 180° turns at the two thymine bases. The folded structure is stabilized by stacking interactions, hydrogen bonds and ionic interactions. Molecular simulations using this double-folded structure as a template for modeling trinucleotide repeats suggest that the folded motif is reasonably stable, which may affect DNA transcription and/or subsequent RNA translation.

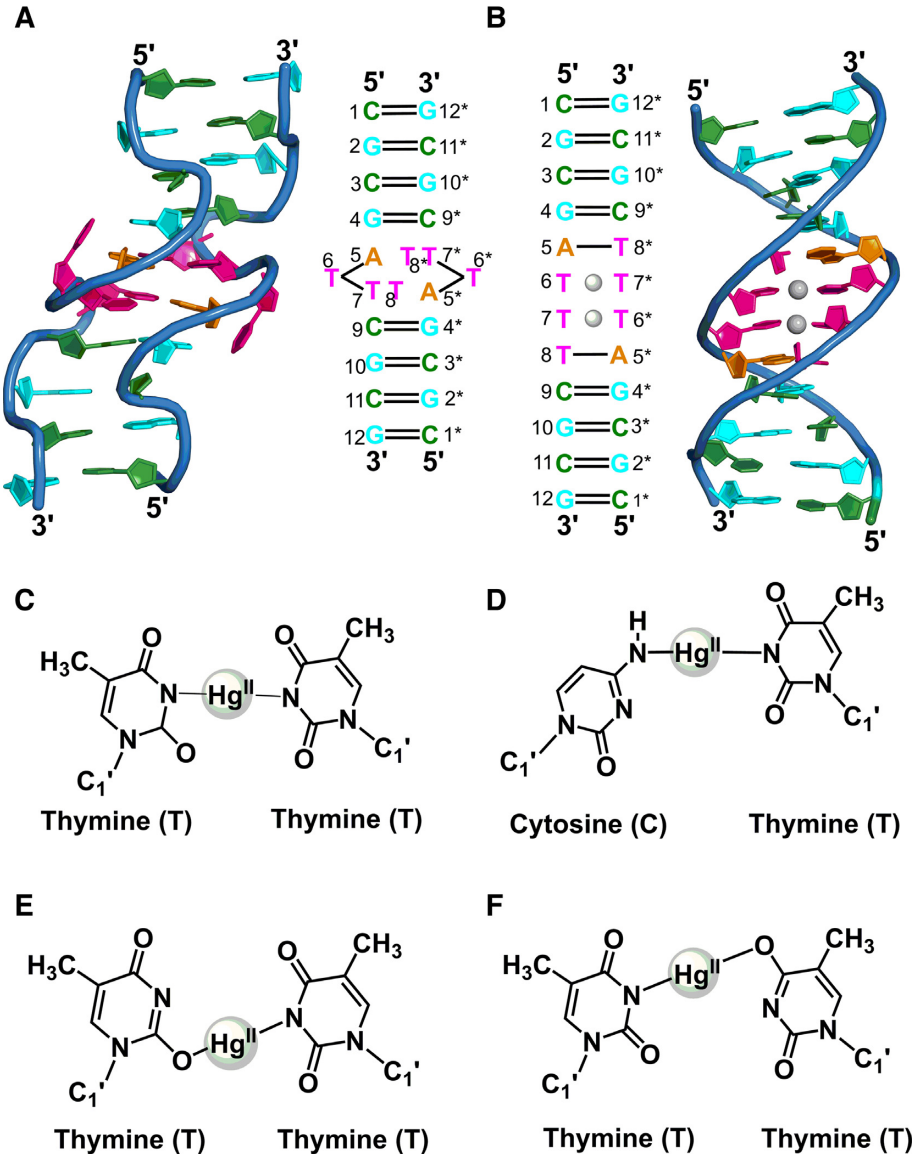

Figure 2. Schematic representations and overall structures of (**A**) T:T DNA duplex (PDB ID: 4L25) and (**B**) T–Hg^{II}–T complex (PDB ID: 4L24). Hg^{II}-mediated interactions are shown in (**C**) for the N3–Hg^{II}–N3 interaction between the T:T pair, (**D**) for the C–Hg^{II}–T interaction, (**E**) for the O2–Hg^{II}–N3 interaction between the T:T pair and (**F**) for the N3–Hg^{II}–O4 interaction between the T:T pair.

Conformational dynamics may also play an important role in repeat expansion. Recently, Huang *et al.* showed that TGGAA repeats, which are involved in spinocerebellar ataxia type 31, exhibited hairpin structure polymorphism (64). The end-to-end aligned configuration was favored when the DNA contained odd number of repeats, whilst the overhang configuration was favored when the number of repeats was an even number. For large repeat numbers the end-to-end configuration became dominant. Structural analysis by X-ray crystallography showed that the GGA motifs within the stem region of the hairpin are kinked and may act as hot spots to facilitate the transition between the two configurations (Figure 4B). Based on their findings, the authors propose a consecutive expansion model for TGGAA tandem repeats where the DNA initially slips between the two configurations but eventually settles on the expansion-promoting end-to-end configuration in longer repeats. This model may also be applicable to other repetitive DNA systems.

The dumbbell and mini-dumbbell (MDB) are new types of native DNA structure which may be among the structural intermediates involved in repeat expansion processes (65,66) (Figure 4C). MDBs comprise two tetranucleotide type II loops where the first and fourth loop residues form a loop-closing base pair, whilst the second and third residues fold into the minor groove and stack onto the loop-closing base pair. MDBs are highly compact structures stabilized by extensive loop–loop interactions. However, MDBs may 'melt' and exchange with the miniloop conformation, leading to repeat expansion. The ability to exchange between the MDB and miniloop structures may also provide a potential pathway to escape recognition from DNA repair proteins.

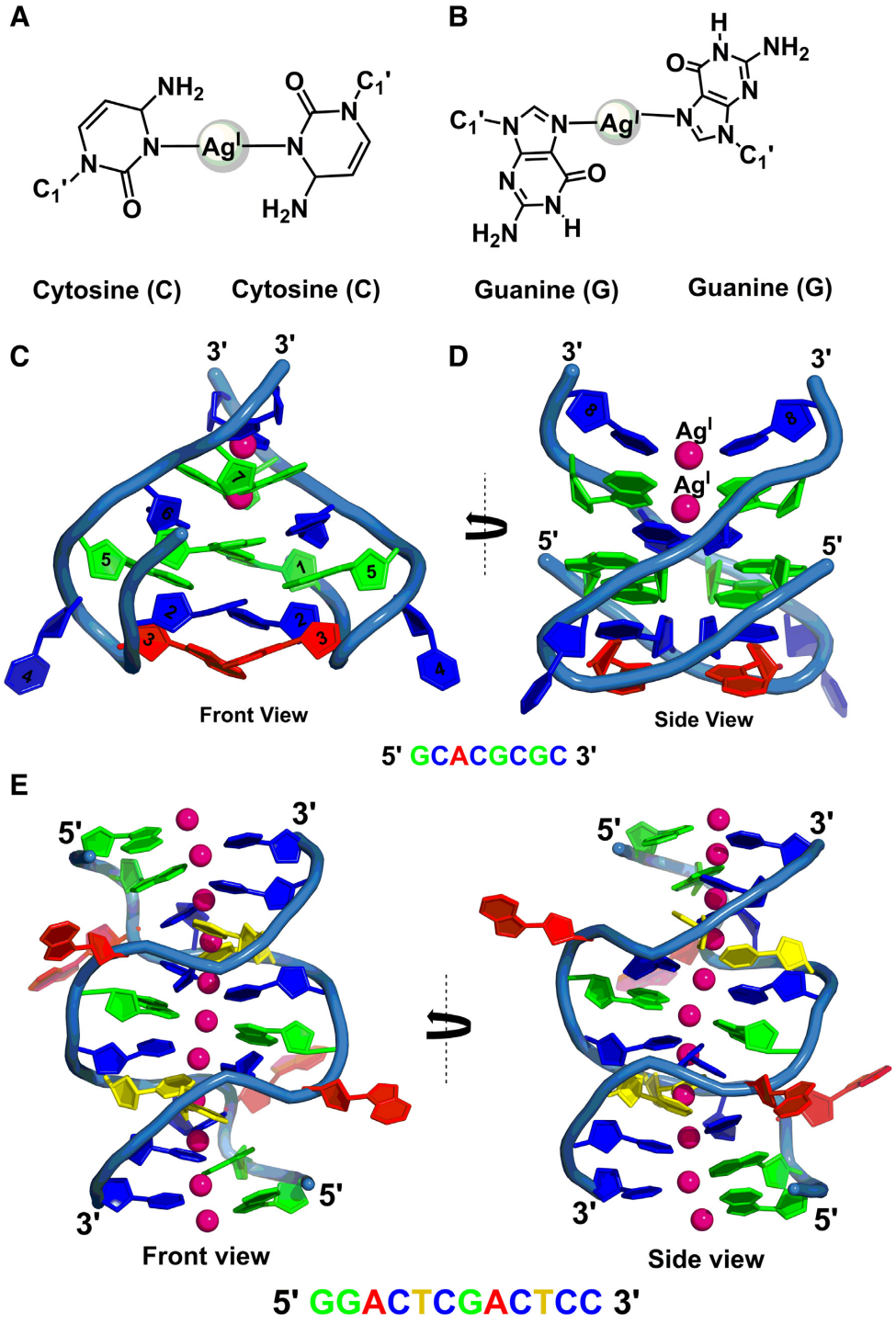

Figure 3. Schematic view showing the detailed (**A**) C–Ag^{I}–C and (**B**) G–Ag^{I}–G interactions. The sequence of the DNA and the structure of the DNA–Ag^{I} complex (PDB ID: 5XJZ) are shown in (**C**) and (**D**) The DNA strands are shown as cartoon models with the silver ions represented as hot-pink spheres. Numbering of the nucleotides is from 5′ to 3′ end of the DNA. (**E**) Crystal structure of metallo-DNA nanowire dodecamer duplex containing Ag-mediated mismatch base pairs view from front and side. One-dimensional arrays of silver ion along the DNA helical axis are shown in hot pink spheres (PDB ID: 5IX7).

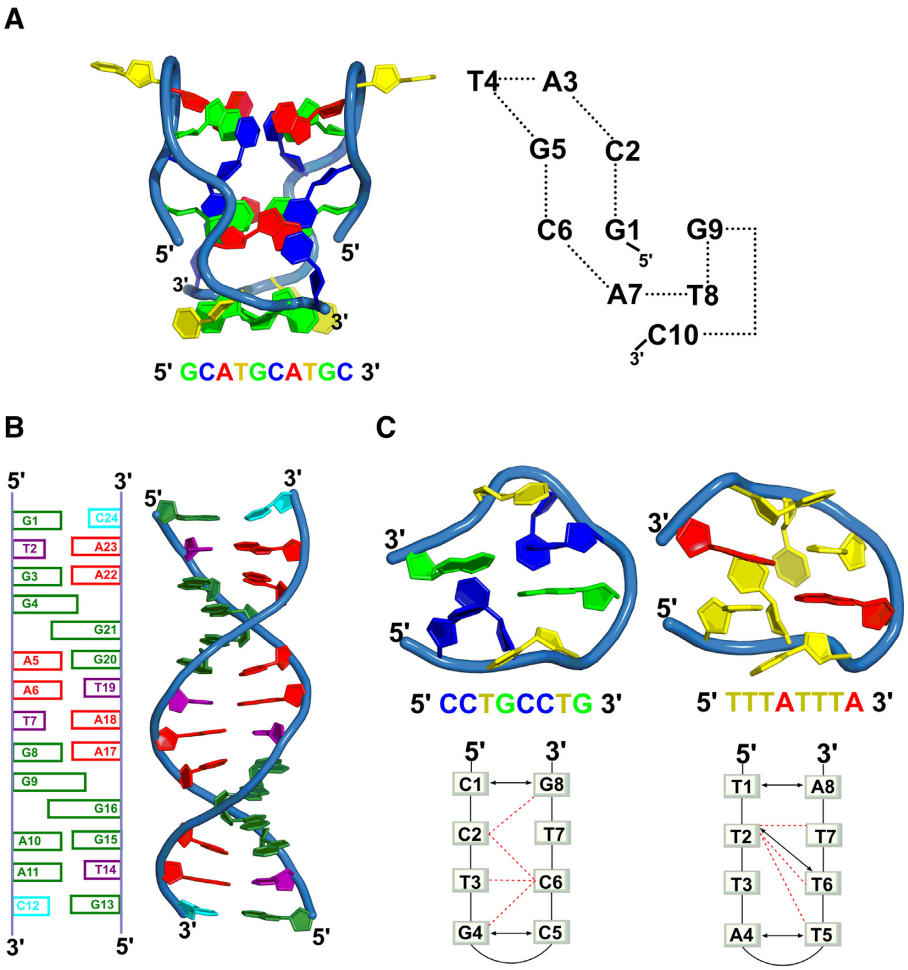

Figure 4. (**A**) Crystal structure of d(GCATGCATGC) double folded structure of the dodecamer (PDB ID: 4ZKK) viewed from front (left). A schematic representation of the stacking and the connectivity of the dodecamer sequence is shown on the right. (**B**) Structure of the $dG(TGGAA)_2C$ duplex (PDB ID: 5GUN). A schematic of the base–base interactions between the two strands is shown on the left, and the cartoon representation of the duplex structure is shown on the right. (**C**) Mini-dumbbell structures and loop–loop interactions in CCTG (left) (PDB ID: 5GWL) and TTTA (right) (PDB ID: 5GWQ). The stacking and hydrogen bond interactions are represented as black arrows and red dotted lines, respectively.

The above studies are additional examples of the structural versatility of the DNA molecule and further attests to the viewpoint that the ability to interchange between different structures may also be important in its functional mechanism. Although none of the proposed models above have been established *in vivo* due to their recentness, they do provide testable frameworks for studying the mechanism of repeat expansions.

LIGAND-INDUCED STRUCTURES OF THE DNA DUPLEX

Arguably most progress in the field of unusual DNA structures during the past decade has been in the area of DNA–small ligand complexes. The studies of DNA–small ligand complex structures are especially important if one intends to develop therapeutic or diagnostic molecules that target the DNA. Interestingly, many of the structural features observed for small ligand–DNA complexes have also been reported in DNA–protein complexes, suggesting that these two may share similar interaction mechanisms (67–72) (Figure 5). Thus the study of DNA–small ligand complex structures may also improve our understanding of how proteins target DNA (73). Recent structures can be assigned to two broad categories: Those causing severe helix bending whilst retaining the base pairing pattern (Table 2) and those causing nucleotide flip-out (Table 3). We shall describe some of the more interesting structures in the following sub-sections.

DNA bending induced by organic compounds

Chromomycin A_3 (Chro) (Figure 6A) and mithramycin (Mith) (Figure 6B) are aureolic acid drugs with anti-tumor activity which bind to GC-rich DNA sequences (74,75). Their anti-tumor activity has been ascribed to the inhibitory effect of these compounds on DNA replication and RNA transcription. Both compounds form dimers which are mediated by divalent metal ions when binding to the DNA (76). The crystal structure of a Chro–$d(TTGGCCAA)_2$ complex showed that the Mg(II)-chelated Chro dimer binds to the minor groove of the GGCC sequence and significantly widens the groove in the process (77) (Figure 6C).

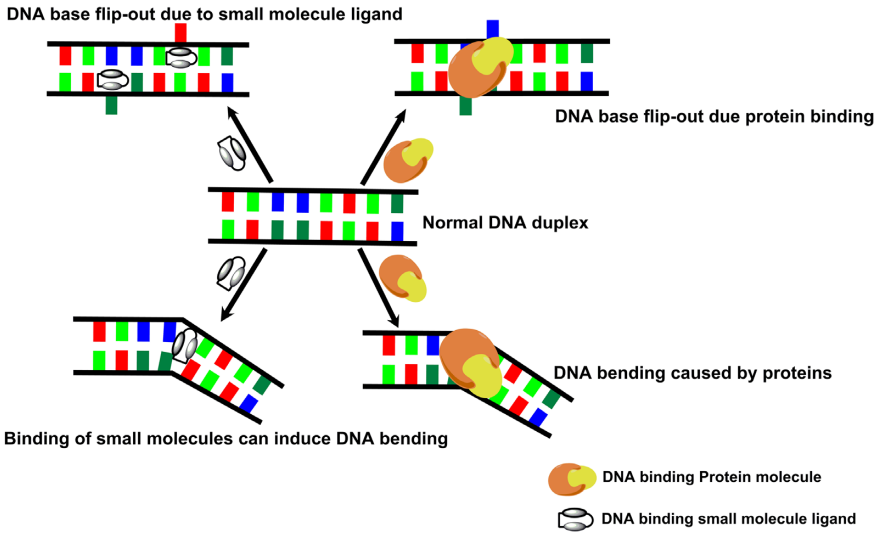

Figure 5. Schematic representation of how small molecule ligands and/or proteins may induce DNA bending or base flip-out in DNA duplexes.

The metal ion is coordinated to the O1 and O9 oxygen atoms of each chromophore within the dimer and forms an octahedral coordination system with two water molecules acting as the fifth and sixth ligands. Hydrogen bonds between Chro and the guanine bases may explain the preference for GC-rich bases. As a consequence, the DNA is kinked by ∼33° with the TpG steps showing a positive roll of ∼17° in the major groove. The structure of a Mith analogue, Mith SA-Phe (Figure 6B), in complex with d(AGGGTACCCT)$_2$ DNA revealed that the small compound is capable of binding to 5′-AGGG and 5′-GGGT in the Zn(II)–chelated dimer form (78) (Figure 6D). Similar to Chro, the Zn ion is octahedrally coordinated to the oxygen atoms of the chromophore and two water molecules, and binding of Mith SA-Phe to the DNA introduces kinking and bending of the DNA duplex.

Even well studied interactions between DNA and small molecules may harbor new surprises. Echinomycin (Echi), a quinoxaline antibiotic from *Streptomycetes* is a good example. Echinomycin is a bis-intercalator which binds to the DNA duplex and induces bending at CpG sites (79). The structural bases behind this interaction have been extensively studied (80–82). However, recently our group found that echinomycin prefers to bind to consecutive CpG steps separated by a T:T mismatch rather than perfect DNA duplexes (Figure 7G) (83). Detailed structural analysis showed that echinomycin induces deformation in the mismatched DNA duplex, allowing cooperative recognition of the T:T mismatch by a second echinomycin molecule. Sequences harboring multiple CAG/CTG repeats are prone to form intra-strand T:T mismatches and have been associated with increased cancer risk in the past (84,85). This recent finding may be useful in re-assessing the utility of echinomycin for detection of MMR-deficient cancers, possibly by leveraging the DNA binding-induced fluorescence quenching properties of echinomycin.

Ruthenium polypyridyl compounds (Figure 7A–D) may also introduce kinks into the DNA duplex structure. These compounds have received wide attention due to their increased photoemission upon binding to DNA, especially mismatch DNA sequences (86). This 'light switch' effect may thus have the potential to become a diagnostic tool (87). The structure of Λ-[Ru(TAP)$_2$(dppz)]$^{2+}$ (TAP = tetraazaphenanthrene, dppz = dipyridophenazine) complexed with d(TCGGCGCCGA)$_2$ revealed that the dppz group intercalated into the TC steps at both ends of the duplex, whilst one of the TAP moieties semi-intercalated into the GG steps of a second, symmetrically equivalent DNA duplex (88) (Figure 7E). Disruption of the GG stacking introduces a 51° kinked bending that was stabilized by a Ba(II) ion. The CG steps are also bent by 22°. Interestingly, the structure of the complex changes from an A-/B-DNA hybrid conformation to the A-DNA conformation when the relative humidity is lowered from ∼97% to 84–79%, accompanied by an increase in the kink angle at the central CG step from 22° to 51° (89) (Figure 7F). This transition is reversible upon rehydration. The complex structures of Λ- and Δ-[Ru(phen)$_2$(dppz)] (phen = phenanthroline) enantiomers bound to d(TCGGCGCCGA)$_2$ and d(TCGGCGCCGA)$_2$, respectively, revealed that both compounds also induced DNA kinking via semi-intercalation binding modes by inserting the phen groups into the GG steps of the DNA duplexes. However, the intercalation angles between the dppz groups and the terminal GG steps are different from the Λ-[Ru(TAP)$_2$(dppz)]$^{2+}$ structure, suggesting that the different organic moieties neighboring the same organic group may affect its mode of binding.

Allosteric structural perturbations by pyrrole-imidazole polyamides

Pyrrole (Py)-imidazole (Im) polyamides (PIPs) forming hairpin-shaped oligomers can bind to a broad repertoire of DNA sequences (90). Hairpin PIPs bind specifically to the minor groove of DNA sequences with the following rules: parallel and antiparallel Py/Im pairs recognize C:G and G:C base pairs, respectively, whilst Py/Py pairs recognize either A:T or T:A base pairs (91–93). The crystal structure of an 8-ring cyclic PIP complexed with d(CCAGGCCTGG)$_2$

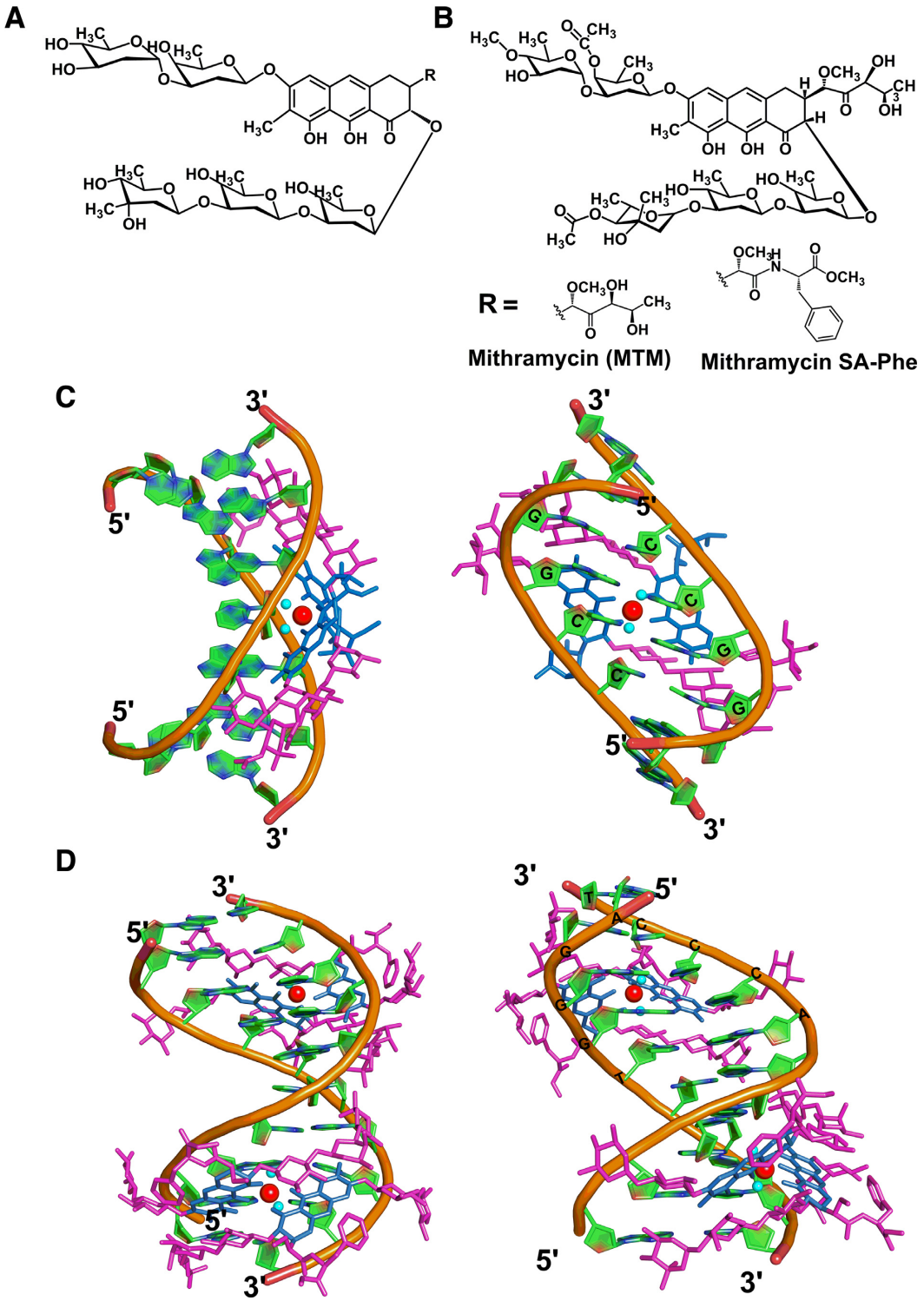

Figure 6. Chemical structures of (**A**) chromomycin A3 and (**B**) mithramycin along with its analogue. (**C**) Overall structure of the Mg^{II}-coordinated Chro-$(TTGGCCAA)_2$ complex viewed from the side (left) and GGCC binding sites through the major groove (right). The DNA backbone is coloured in orange, di- and tri-saccharides in light magenta and chromophore in skyblue, Mg^{II} in red and the two water molecules in cyan. (**D**) Overall structure of the Zn^{II}-coordinated Mith SA-Phe-(AGGGTACCCT)2 complex viewed from the minor groove (left) and 5′-AGGG binding site through major groove (right). The DNA backbone is coloured in orange, the Mith SA-Phe saccharides in light magenta and chromophore in skyblue, Zn^{II} in red and the two water molecules in cyan.

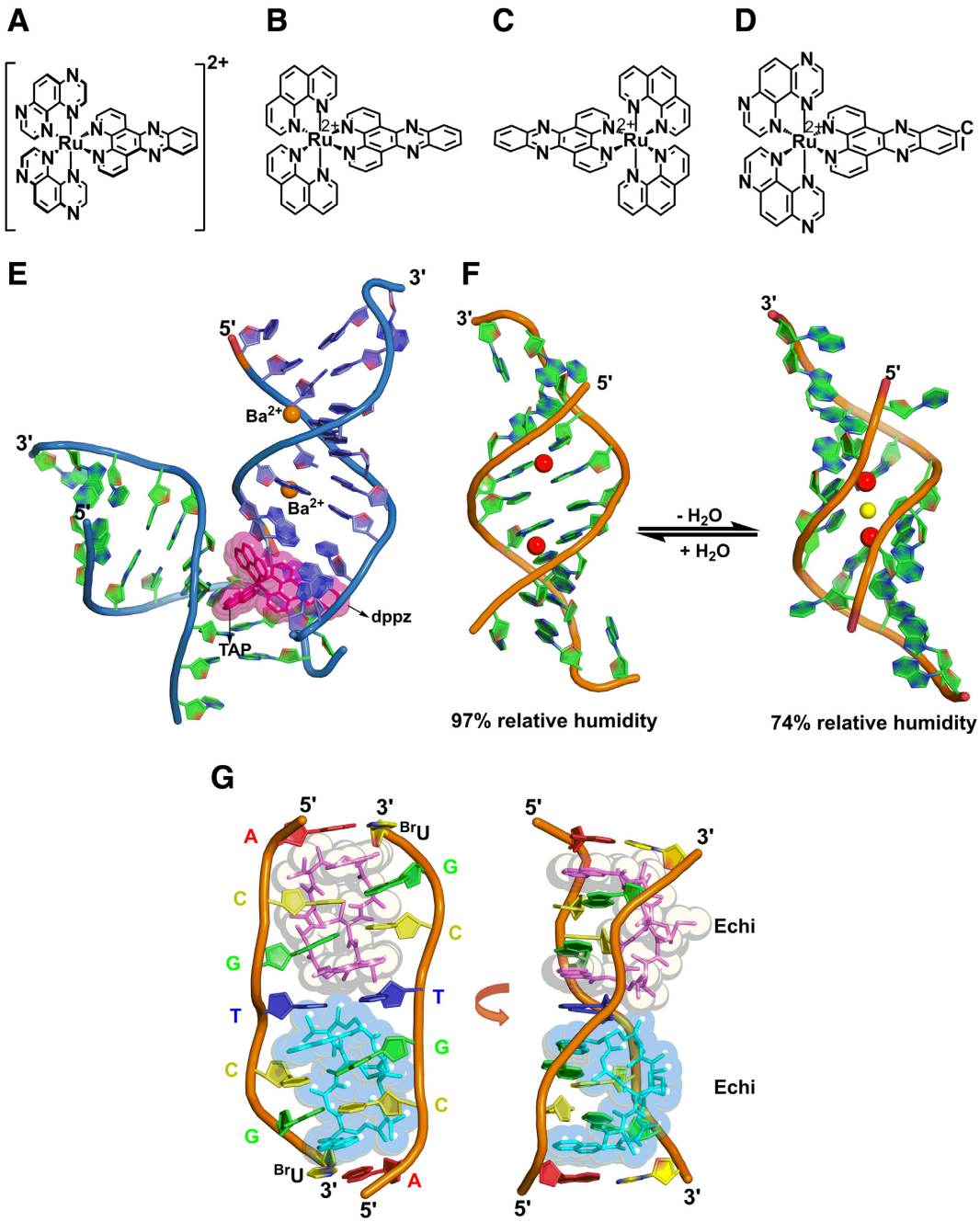

Figure 7. Chemical structures of (**A**) Λ-$[Ru(tap)_2(dppz)]^{2+}$, (**B**) Λ-$[Ru(phen)_2dppz]^{2+}$, (**C**) Δ-$[Ru(phen)_2(dppz]^{2+}$, (**D**) Λ-$[Ru(TAP)_2(11\text{-Cl-dppz})]^{2+}$. (**E**) Λ-$[Ru(TAP)_2(dppz)]^{2+}$ exhibits both semi-intercalation and intercalation binding modes when complexed to $d(TCGGCGCCGA)_2$. (**F**) The Λ-$[Ru(TAP)_2(dppz)]^{2+}$ -DNA complex of (E) undergoes structural changes when subjected to different humidity conditions. Left: structure at 97% relative humidity (PDB ID: 4LTF); Right: structure at 74% relative humidity (PDB ID: 4LTG). (**G**) Crystal structure of the echinomycin–TT duplex (PDB: 5YTZ). The asymmetric unit of the echinomycin–TT complex structure in front (left) and side (right) views. Two echinomycin molecules bind to a duplex DNA containing one T:T mismatch.

has been solved (94) (Figure 8A). The cyclic PIP contains two antiparallel Im-Im-Py-Py strands capped by (*R*)-α-amino-γ turn units at both ends and binds to the central six base pairs of the DNA duplex. The complex structure revealed a 4 Å widening of the minor groove and compression of the major groove along with a >18° bend in the helix axis towards the major groove. The (*R*)-α-amino-γ turns recognize the A:T base pairs of the duplex by forming hydrogen bonds among the amides, imidazoles and water molecules. Another cyclic PIP containing β-amino-γ turns instead of α-amino-γ turns was found to bind to the central six base pairs of the sequence $d(CCAGTACTGG)_2$ and modulate DNA shape (95). Interestingly, the structural alteration caused by these minor groove-binding synthetic cyclic PIPs disrupted the binding of major groove-binding transcription factors such as the androgen and glucocorti-

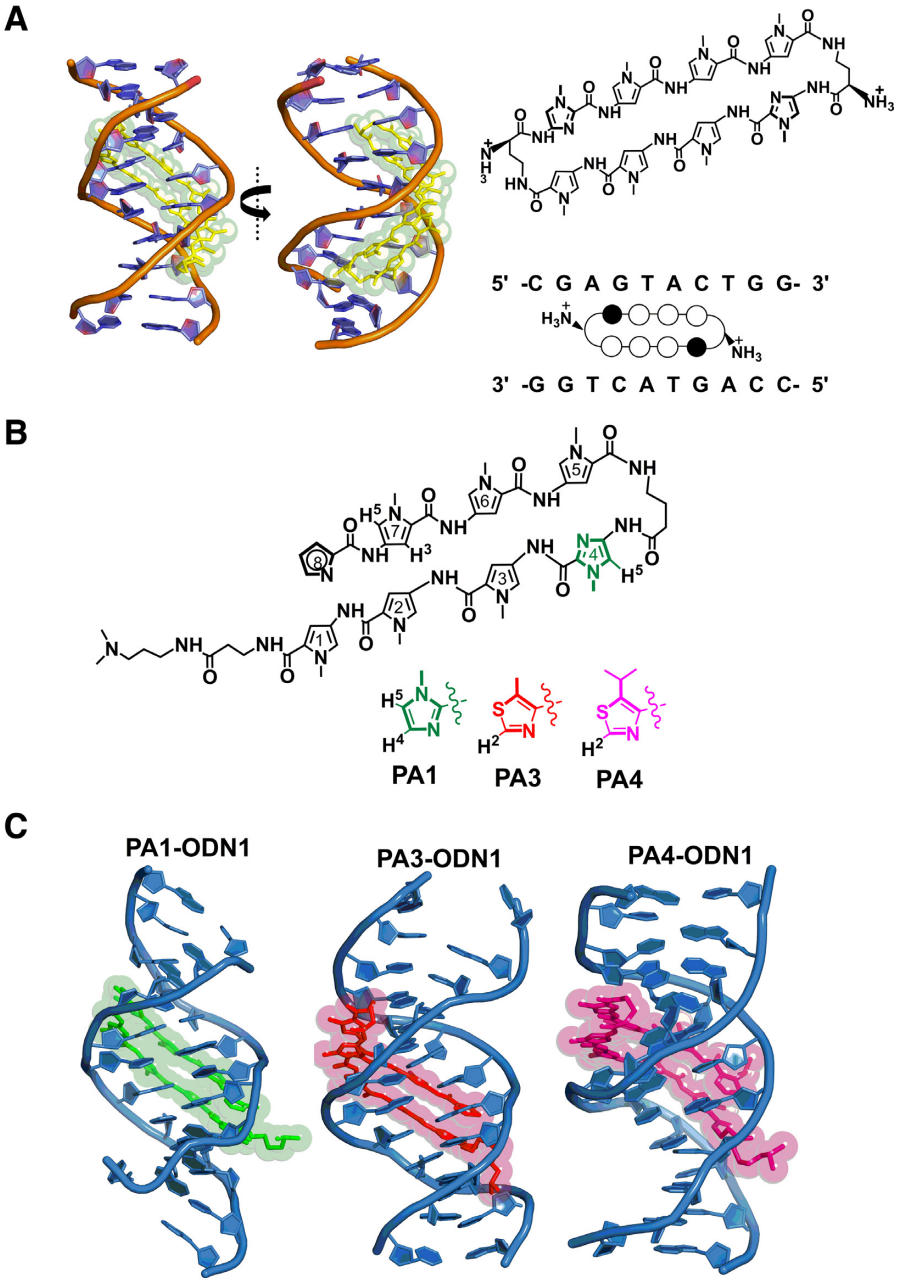

Figure 8. (**A**) The crystal structure of an 8-ring cyclic Py/Im polyamide in complex with d(CCAGGCCTGG)$_2$ viewed from the front and sideways (left). The chemical structure of the cyclic polyamide and the sequence of the DNA are shown on the right. Black circles represent imidazoles, open circles represent pyrroles, and ammonium substituted half circles at each end represent the (*R*)-amine-turn. (**B**) Chemical structures of the 5-alkyl thiazole-substituted PIPs PA1, PA3 and PA4. (**C**) Crystal structures of PA1•ODN1 (PDB ID: 5ODF), PA3•ODN1 (PDB ID: 5ODM) and PA4•ODN1 (PDB ID: 5OE1) complexes viewed from the major groove.

coid receptors (96). This allosteric perturbation of the DNA helix provides a molecular basis for disruption of transcription factor–DNA interfaces by small molecules and may be applied to the chemical control of gene networks (97–100).

Padroni et al expanded the heterocyclic repertoire of hairpin PIPs by replacing the N-terminal Im unit of a linear PIP with different 5-alkyl thiazole (Nt) building blocks (101) (Figure 8B). They demonstrated that the different Nt building blocks differed in their relative binding preferences for a target guanosine in the minor groove. The selectivity of the Nt for guanosine increased with increasing steric bulk at the 5-position of the Nt. NMR-derived models showed that hairpin PIPs which incorporate an N-terminal Nt unit induced a more pronounced compression of the major groove of target DNA duplexes compared to archetypical hairpin PIPs (Figure 8C). On the other hand, the helix bending caused by Nt-containing PIPs is reduced compared to archetypical PIPs. The availability of different PIPs which induce different structural changes in DNA could offer new

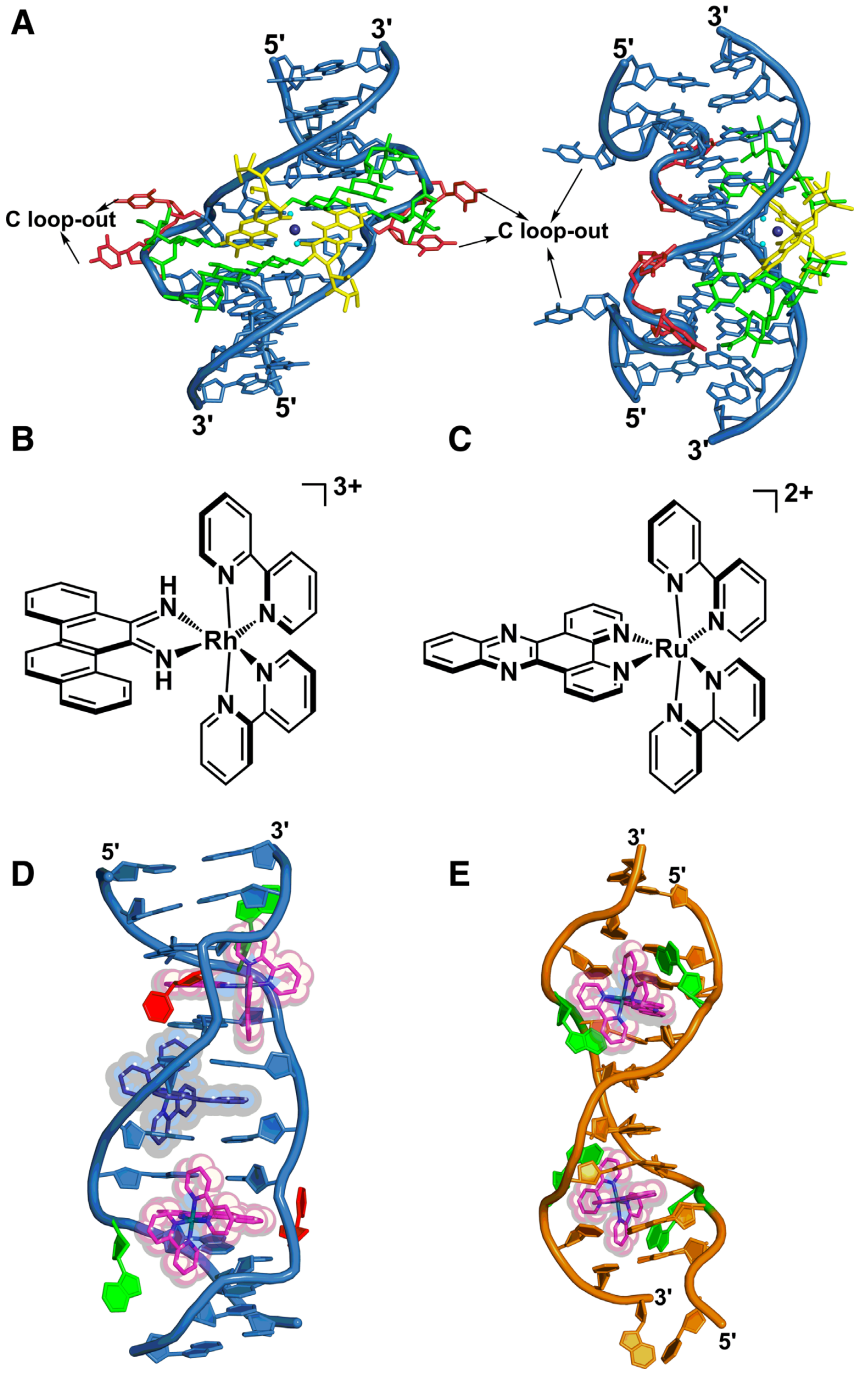

Figure 9. (**A**) Crystal structure of the $Ni^{II}(Chro)_2$-d[TT$(CCG)_3$AA]$_2$ complex viewed from the minor groove (PDB ID: 5XEW). The DNA backbone is coloured in light blue and the flipped-out cytocines are represented in red. The chromomycin A3 dimer chromophores are shown as green and yellow sticks whilst the Ni^{2+} is shown as a dark blue sphere in coordination with two water molecules, shown in cyan. (**B**) Chemical structure of Δ-$[Rh(bpy)_2chrysi]^{3+}$. (**C**) Chemical structure of Δ-$[Ru(bpy)_2dppz]^{2+}$. Overall structure of the rhodium complex bound to the oligonucleotide 5′-CGGAAATTCCCG-3′. (**D**) Insertion of Δ-$[Rh(bpy)_2chrysi]^{3+}$ (red) into d$(CGGAAATTCCCG)_2$ (PDB ID: 2O1I) occurs via the minor groove and extrudes the A:C mismatch. The ejected adenosine (green) remains in the minor groove, whereas the ejected cytosine (red) is now located in the major groove. (**E**) Structure of Δ-$[Ru(bpy)_2(dppz)]^{2+}$ bound to d$(CGGAAATTACCG)_2$ (PDB ID: 4E1U). Two Δ-$Ru(bpy)_2(dppz)^{3+}$ molecules (pink) are inserted into the two A:A mismatch sites. The ejected adenosines are colored in green.

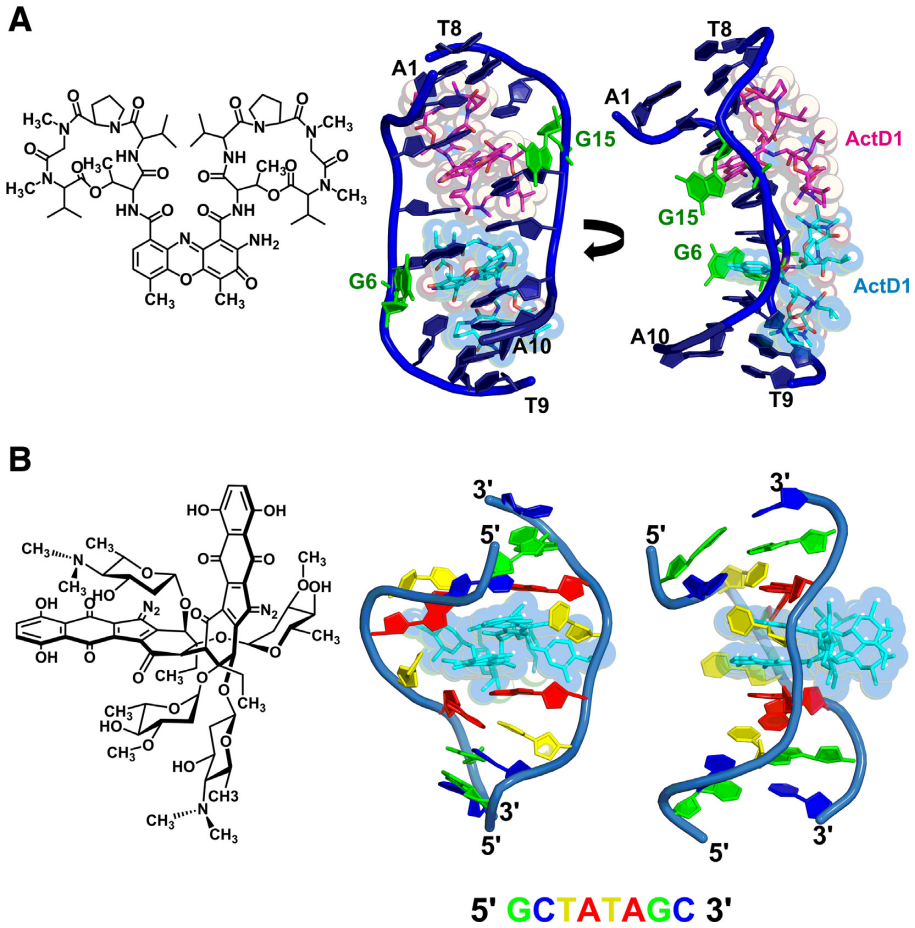

Figure 10. (**A**) Chemical structure of actinomycin D (left) and crystal structure of the 2:1 ActD–$(ATGCGGCAT)_2$ DNA complex (PDB ID: 4HIV) (right). Two ActD molecules, ActD1 (pink) and ActD2 (cyan), bind to DNA by intercalating their phenoxazone rings at the two GpC steps. The cyclic pentapeptide moieties are located in the minor groove. Two guanines, G6 and G15, (colored in green) are looped out with the base plane perpendicular to the long axis of the flanking G:C base pairs. (**B**) Chemical structure of (–)-lomaiviticin A (left) and solution structure of (–)-lomaiviticin A complexed to $d(GCTATAGC)_2$ (right) viewed from the front and from the side (PDB ID: 2N96).

opportunities for controlled modulation of gene expression through allostery.

Metal ion-containing compounds may cause nucleotide flip-out

Several studies have shown that small-molecule ligand may induce nucleotide flip-out of mismatched base pairs within the hairpin stem of trinucleotide repeats (TNRs) (102). Since TNR expansions are associated with the pathogenesis of various neurological diseases, the structural basis behind the ability to specifically recognize these mismatches is of considerable interest to both basic and applied researchers. Recently, recognition of nickel-chelated chromomycin dimer ($Ni^{II}(Chro)_2$) to CCG TNRs has been shown to employ a classic induced-fit paradigm (103). Binding of the compound to the DNA induces large-scale deformations, including the extrusion, or flip-out, of two consecutive cytosines (Figure 9A). This allows the rest of the bases to form a GGCC tetranucleotide core in the minor groove of the DNA duplex. The overall structure of the GGCC binding core in the $Ni^{II}(Chro)_2$-$d[TT(CCG)_3AA]_2$ complex is similar to the GGCC tetranucleotide region in a previously reported $Mg^{II}(Chro)_2$-$d[TTGGCCAA]_2$ complex (77), which does not contain any nucleotide extrusions. This suggests that the GGCC tetranucleotide core is the cognate structure for metal-chelated chromomycin dimers, and the compound may be able to bind to the DNA as long as it is able to 'force' the DNA duplex to assume this core conformation.

Rhodium and ruthenium complexes that recognize mismatched sites with high selectivity may also induce nucleotide flip-out in addition to more moderate structural perturbations such as kinking. The structure of Δ-$Rh(bpy)_2(chrysi)^{3+}$ (bpy = 2,2′-bipyridine, chrysi = 5,6-chrysenequinone diimine) (Figure 9B) bound to a palindromic duplex containing two C:A mismatches has been determined (104) (Figure 9D). The structure showed that Δ-$Rh(bpy)_2(chrysi)^{3+}$ intercalates into the major groove of matched base pairs in the DNA with the chrysi ligand interacting shallowly at the major groove. In contrast, the chrysi ligand is deeply inserted into the mismatch site, which is located in the minor groove, and extrudes the mismatched C:A base pair into the major groove while forming π-stack interactions with neighboring base pairs. The

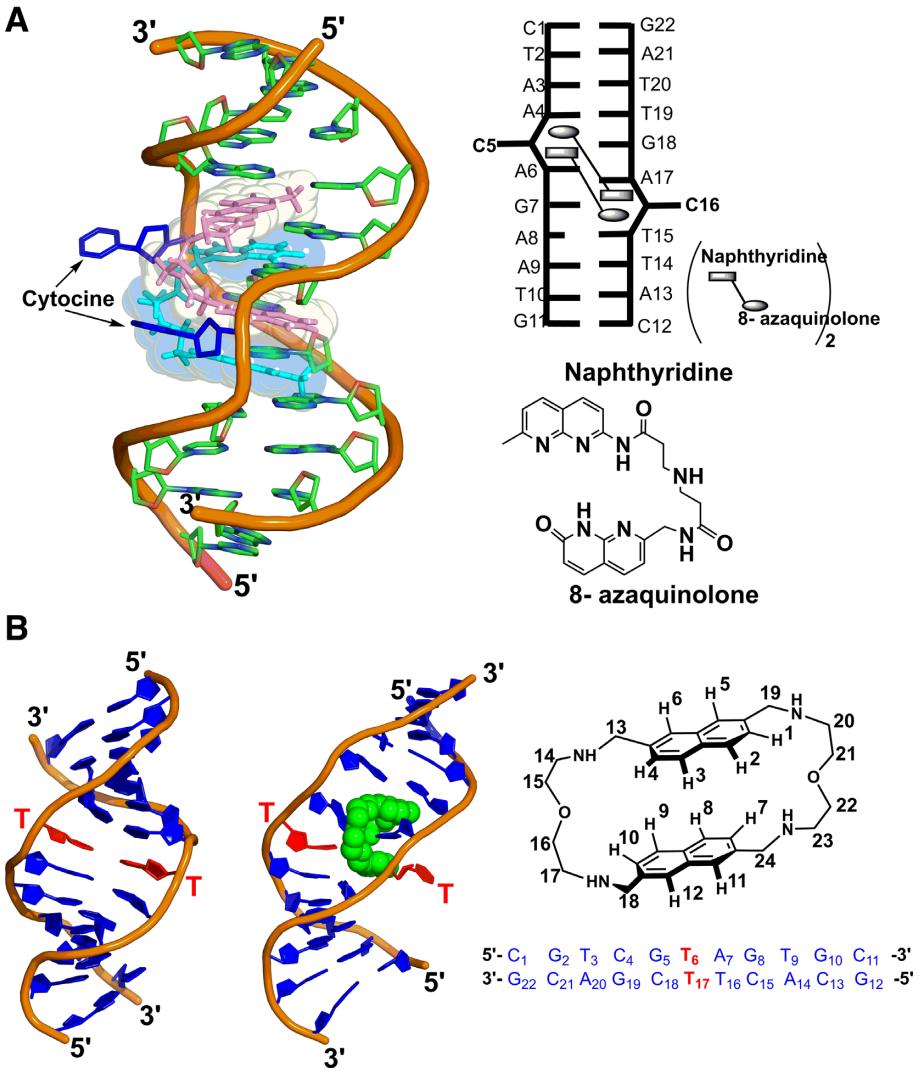

Figure 11. (**A**) Solution structure of the NA–CAG–CAG complex (PDB ID: 1x26) (left). The extruded cytosines are coloured in blue, and the two subunits of the NA dimer are coloured in light pink and cyan. A schematic representation of the conformational changes induced by the binding of NA to DNA is shown on the top right panel. The chemical structure of NA is shown on the lower right panel. (**B**) Structures of T–T DNA duplex in the absence (PDB ID: 2LL9) and presence (PDB ID: 2LLJ) of 2,7-BisNP (green spheres) (left). The ligand binds to the DNA through a threading intercalation mode. The DNA backbone is shown as an orange cartoon, with the thymine bases coloured in red. The chemical structure of 2,7-BisNP and the sequence of the TT–DNA duplex are shown on the right. The T:T mismatch is highlighted in red.

mismatched cytosine is extruded into the major groove, where it is positioned perpendicular to the base pairing plane of the helix. The ejected adenosine remains in the minor groove. Subsequent NMR and crystallographic studies of Δ-Rh(bpy)$_2$(chrysi)$^{3+}$ bound to oligonucleotides containing other C:C or A:A mismatches confirmed the presence of the same insertion modes (105,106). The ruthenium complex Ru(bpy)$_2$dppz^{2+} (Figure 9C) binds to A:A mismatches by inserting the dppz ligand into the mismatched site in the minor groove (Figure 9E) (107). The mismatched adenosines are extruded, with the dppz effectively taking their place and forming stacking interactions with the flanking base pairs. The ejected adenosines are folded back in the minor groove. Both rhodium and ruthenium complexes have photochemical activities that may be desirable in biomedical applications. Rhodium complexes may cleave DNA at the binding site upon photoactivation, and ruthenium complexes may be used to detect the presence of DNA mismatches.

Nucleotide flip-out caused by organic compounds

A number of organic compounds may also cause large structural deformations and induce nucleotide flip-out in DNA. The venerable anti-cancer compound actinomycin D (ActD) (Figure 10A) has high affinity towards neighbouring GpC sites flanking a G:G mismatch, which may have applications in TNR expansion detection and therapy. The structure of ActD in complex with d(ATGCGGCAT)$_2$, which contains two GpC sites separated by a G:G mismatch, revealed that two ActD molecules bound to adjacent overlapping GpC sites, with two guanines looped out and located perpendicular to the long axis of the flanking G:C base pairs (108) (Figure 10A). The binding also

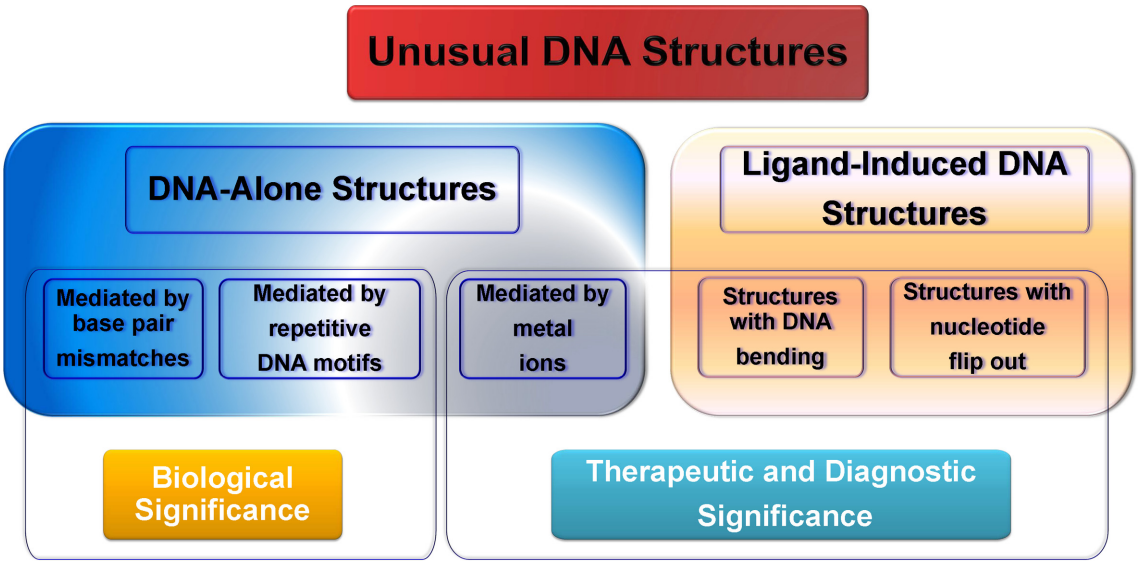

Figure 12. Schematic diagram summarizing unusual DNA structures provoked by mismatches, repeats, ligand binding in the review.

causes a sharp bend and a left-handed twist of the DNA helix. (–)-lomaiviticin A is another natural compound with anticancer activity that induces flip-out of nucleotides. (–)-lomaiviticin A possesses a C_2-symmetric structure that contains two unusual diazofluorene groups (109) (Figure 10B). Nucleophilic activation of each diazofluorene within (–)-lomaiviticin A produces vinyl radical intermediates that induces DNA double-strand breaks. The structures of (–)-lomaiviticin A bound to DNA have been determined by NMR spectroscopy and computational modeling (Figure 10B). These structures show that both diazofluorene residues of (–)-lomaiviticin A penetrate the DNA duplex and disrupt normal base pairing, which leads to ejection of the central AT bases. The reactive centers of (–)-lomaiviticin A are located in close proximity to each strand in the process, thus explaining the DNA cleavage activity of the ligand complex.

Naphthyridine analogues may also cause nucleotide extrusion of DNA. Nakatani's group developed a naphthyridine-azaquinoline (NA) ligand with high affinity towards the CAG/CAG triad, which contains an A:A mismatch (102). The structure of the NA-CAG/CAG complex revealed that NA bound to the mismatched duplex as a dimer, with the two mismatched adenosine bases forming intermolecular hydrogen bonds with the 8-azaquinolone units of the two NA molecules. The most striking feature of the NA-CAG/CAG structure is the invasion of the G:C base pairing sites by naphthyridine moieties, which forces the cytidine nucleotides to extrude out and form two C-bulge loops (Figure 11A).

The bis-naphthalene macrocycle (2,7-BisNP) is an example of a cyclic bis-intercalator which recognizes T:T mismatch sites in duplex DNA with high affinity and selectivity (110). NMR-restrained molecular modeling of 2,7-BisNP bound to d(CGATCGTAGTGC)/d(GCACTTCGACG) duplex revealed that one of the naphthalene units may occupy the space normally occupied by one of the mismatched thymines, resulting in its extrusion (Figure 11B). The second naphthalene unit may intercalate at the A:T base pair flanking the mismatch site, leading to encapsulation of the base-paired thymine residue via double stacking. The encapsulated thymine forms hydrogen bonds with the polyammonium linkers of 2,7-BisNP, which are located at both the minor and major grooves of the oligonucleotide. The remaining mismatched thymine within the helix also forms hydrogen bonds with the linkers. The hydrogen bonds associated with the linkers may help stabilize the ligand–DNA complex.

PERSPECTIVE AND CONCLUSION

There can be no doubt that DNA duplexes are remarkably flexible and able to assume a variety of structures. This flexibility is inherent in the function of DNA as our genetic material. The classical A-, B- and Z-form DNA are all linked to biological processes such as protection against desiccation in bacteria (A-DNA) (111), normal duplex formation (B-DNA) and transcription rate regulation (Z-DNA) (112). Repetitive DNA sequences may assume various conformations which could be associated with pathological repeat expansion and escape from DNA-repair proteins (see references listed in Table 1). Ligands that bind to the DNA duplex induce various degrees of structural deformations, with many of them having pharmaceutical potential (Figure 12).

Given the variety of static structures available, two linked questions arise: (i) which structures are accessible to the DNA under a biological context, and (ii) how the DNA dynamically interchanges between these different structures. It is important to realize that structures which were only observed under certain conditions within the laboratory may also exist transiently under physiological conditions and thus play a role in biological processes. A classic example is Z-DNA. Although it was discovered in 1979, it took almost a decade to discern its first biological role, and the various activities of Z-DNA in life are still not fully understood even to this day (113,114). It is possible (and quite probable) that we are looking at years of studies ahead before we may

glimpse the biological implications of many of the more recent unusual DNA structures. Conversely, one should keep in mind that not all reported structures may be accessible to the DNA molecule within living cells. Recent advances in instrumentation and methodology have started to provide researchers with a handle to detect and characterize transient structures. For example, single-molecule fluorescence experiments can detect low-abundance conformations which interchange at the millisecond time scale (115), and NMR relaxation dispersion techniques allow the detection and structural modeling of low-abundance or transient structures interchanging at the microsecond–millisecond time scales (116). Increasing accuracy in DNA force fields has also enhanced our ability to interrogate the conformational dynamics of DNA at the atomic level through the use of molecular dynamics simulations (117).

The ability to coerce the DNA duplex to assume a certain shape through sequence design and ligand addition may also have applications in bioengineering. DNA nanomaterials with defined shapes and chemical activities may serve as building blocks for molecular devices such as ion sensors, nanowires and DNA magnets (118,119). These devices have the potential to become important research tools in the biological sciences (120).

Whether the aim is to understand DNA duplex structure at the basic level or to apply this knowledge for biomedical engineering, they both require the analysis of assorted structures, preferably solved experimentally, covering as large a conformational space as possible. The fact that the structures of many unusual DNA duplexes, particularly those involving small ligands, were elucidated only recently implies that we are nowhere near understanding how a DNA duplex may fold. We envision that many more unusual structures of DNA duplexes will be found in the future through the efforts of the scientific community. In the end, we hope that our comprehensive review may provide the impetus for people already in the field to persevere along this path and encourage a new generation to join us to explore the fascinating world of DNA duplex structures.

ACKNOWLEDGEMENTS

The authors of this review article would like to acknowledge the National Synchrotron Radiation Research Center (NSRRC), Taiwan for its continuing support in the field of nucleic acids research.

FUNDING

Ministry of Science and Technology, Taiwan [106-2628-M-005-001-MY3 to M.-H.H.]. Funding for open access charge: Ministry of Science and Technology, Taiwan.
Conflict of interest statement. None declared.

REFERENCES

1. Schindler,S. (2008) Model, theory, and evidence in the discovery of the DNA structure. *Br. J. Philos. Sci.*, **59**, 619–658.
2. Franklin,R.E. and Gosling,R.G. (1953) Molecular configuration in sodium thymonucleate. *Nature*, **171**, 740.
3. Wilkins,M.H.F., Stokes,A.R. and Wilson,H.R. (1953) Molecular structure of nucleic acids: molecular structure of deoxypentose nucleic acids. *Nature*, **171**, 738.
4. Watson,J.D. and Crick,F.H. (1953) Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. *Nature*, **171**, 737–738.
5. Doluca,O., Withers,J.M. and Filichev,V.V. (2013) Molecular engineering of guanine-rich sequences: Z-DNA, DNA triplexes, and G-quadruplexes. *Chem. Rev.*, **113**, 3044–3083.
6. Timsit,Y. (1999) DNA structure and polymerase fidelity. *J. Mol. Biol.*, **293**, 835–853.
7. Jaworski,A., Hsieh,W.T., Blaho,J.A., Larson,J.E. and Wells,R.D. (1987) Left-handed DNA in vivo. *Science*, **238**, 773–777.
8. Zhao,J., Bacolla,A., Wang,G. and Vasquez,K.M. (2010) Non-B DNA structure-induced genetic instability and evolution. *Cell. Mol. Life Sci.: CMLS*, **67**, 43–62.
9. Wells,R.D. (2009) Discovery of the role of non-B DNA structures in mutagenesis and human genomic disorders. *J. Biol. Chem.*, **284**, 8997–9009.
10. Wang,G. and Vasquez,K.M. (2014) Impact of alternative DNA structures on DNA damage, DNA repair, and genetic instability. *DNA Repair (Amst.)*, **19**, 143–151.
11. Renciuk,D., Kypr,J. and Vorlickova,M. (2011) CGG repeats associated with fragile X chromosome form left-handed Z-DNA structure. *Biopolymers*, **95**, 174–181.
12. Bowater,R.P. and Wells,R.D. (2001) The intrinsically unstable life of DNA triplet repeats associated with human hereditary disorders. *Prog. Nucleic Acid Res. Mol. Biol.*, **66**, 159–202.
13. Bikard,D., Loot,C., Baharoglu,Z. and Mazel,D. (2010) Folded DNA in action: hairpin formation and biological functions in prokaryotes. *Microbiol. Mol. Biol. Rev.: MMBR*, **74**, 570–588.
14. Huang,C.H., Lin,Y.S., Yang,Y.L., Huang,S.W. and Chen,C.W. (1998) The telomeres of Streptomyces chromosomes contain conserved palindromic sequences with potential to form complex secondary structures. *Mol. Microbiol.*, **28**, 905–916.
15. Astell,C.R., Chow,M.B. and Ward,D.C. (1985) Sequence analysis of the termini of virion and replicative forms of minute virus of mice DNA suggests a modified rolling hairpin model for autonomous parvovirus DNA replication. *J. Virol.*, **54**, 171–177.
16. Kaushik,M., Kaushik,S., Roy,K., Singh,A., Mahendru,S., Kumar,M., Chaudhary,S., Ahmed,S. and Kukreti,S. (2016) A bouquet of DNA structures: Emerging diversity. *Biochem. Biophys. Rep.*, **5**, 388–395.
17. Sheng,J., Gan,J. and Huang,Z. (2013) Structure-based DNA-targeting strategies with small molecule ligands for drug discovery. *Med. Res. Rev.*, **33**, 1119–1173.
18. Gagna,C.E. and Lambert,W.C. (2006) Novel drug discovery and molecular biological methods, via DNA, RNA and protein changes using structure-function transitions: Transitional structural chemogenomics, transitional structural chemoproteomics and novel multi-stranded nucleic acid microarray. *Med. Hypotheses*, **67**, 1099–1114.
19. Guittat,L., Alberti,P., Rosu,F., Van Miert,S., Thetiot,E., Pieters,L., Gabelica,V., De Pauw,E., Ottaviani,A., Riou,J.F *et al.*2003) Interactions of cryptolepine and neocryptolepine with unusual DNA structures. *Biochimie*, **85**, 535–547.
20. Chang,C.K., Jhan,C.R. and Hou,M.H. (2015) The interaction of DNA-binding ligands with trinucleotide-repeat DNA: implications for therapy and diagnosis of neurological disorders. *Curr. Top. Med. Chem.*, **15**, 1398–1408.
21. Boyle,K.M. and Barton,J.K. (2016) Targeting DNA mismatches with rhodium metalloinsertors. *Inorg. Chim. Acta*, **452**, 3–11.
22. Komor,A.C. and Barton,J.K. (2013) The path for metal complexes to a DNA target. *Chem. Commun. (Camb.)*, **49**, 3617–3630.
23. Zalesak,J., Lourdin,M., Krejciota,L., Constant,J.F. and Jourdan,M. (2014) Structure and dynamics of DNA duplexes containing a cluster of mutagenic 8-oxoguanine and abasic site lesions. *J. Mol. Biol.*, **426**, 1524–1538.
24. Fukui,K. (2010) DNA mismatch repair in eukaryotes and bacteria. *J.Nucleic Acids*, **2010**, doi.org/10.4061/2010/260512.
25. Li,G.M. (2008) Mechanisms and functions of DNA mismatch repair. *Cell Res.*, **18**, 85–98.
26. Kunkel,T.A. and Erie,D.A. (2005) DNA mismatch repair. *Annu. Rev. Biochem.*, **74**, 681–710.
27. Kondo,J., Ciengshin,T., Juan,E.C., Sato,Y., Mitomi,K., Shimizu,S. and Takenaka,A. (2006) Crystal structure of d(gcGXGAgc) with X = G: a mutation at X is possible to occur in a base-intercalated

duplex for multiplex formation. *Nucleosides Nucleotides Nucleic Acids*, **25**, 693–704.
28. Kondo,J., Umeda,S., Fujita,K., Sunami,T. and Takenaka,A. (2004) X-ray analyses of d(GCGAXAGC) containing G and T at X: the base-intercalated duplex is still stable even in point mutants at the fifth residue. *J. Synchrotron Radiat.*, **11**, 117–120.
29. Ghosh,A., Kar,R.K., Krishnamoorthy,J., Chatterjee,S. and Bhunia,A. (2014) Double GC:GC mismatch in dsDNA enhances local dynamics retaining the DNA footprint: a high-resolution NMR study. *ChemMedChem*, **9**, 2059–2064.
30. Groothuizen,F.S. and Sixma,T.K. (2016) The conserved molecular machinery in DNA mismatch repair enzyme structures. *DNA Repair (Amst.)*, **38**, 14–23.
31. Rossetti,G., Dans,P.D., Gomez-Pinto,I., Ivani,I., Gonzalez,C. and Orozco,M. (2015) The structural impact of DNA mismatches. *Nucleic Acids Res.*, **43**, 4309–4321.
32. He,G., Kwok,C.K. and Lam,S.L. (2011) Preferential base pairing modes of T.T mismatches. *FEBS Lett.*, **585**, 3953–3958.
33. Gervais,V., Cognet,J.A., Le Bret,M., Sowers,L.C. and Fazakerley,G.V. (1995) Solution structure of two mismatches A.A and T.T in the K-ras gene context by nuclear magnetic resonance and molecular dynamics. *Eur. J. Biochem.*, **228**, 279–290.
34. Muller,J. (2017) Metal-mediated base pairs in parallel-stranded DNA. *Beilstein J. Org. Chem.*, **13**, 2671–2681.
35. Chang,C.C., Lin,L.Y., Zou,X.W., Huang,C.C. and Chan,N.L. (2015) Structural basis of the mercury(II)-mediated conformational switching of the dual-function transcriptional regulator MerR. *Nucleic Acids Res.*, **43**, 7612–7623.
36. Xiao,L., Fu,Z.R., Ding,G.S., Fu,H., Ni,Z.J., Wang,Z.X., Shi,X.M., Guo,W.Y. and Ma,J. (2009) Prediction of survival after liver transplantation for chronic severe hepatitis B based on preoperative prognostic scores: a single center's experience in China. *World J. Surg.*, **33**, 2420–2426.
37. Kuzuya,A. and Ohya,Y. (2014) Nanomechanical molecular devices made of DNA origami. *Acc. Chem. Res.*, **47**, 1742–1749.
38. Kondo,J., Yamada,T., Hirose,C., Okamoto,I., Tanaka,Y. and Ono,A. (2014) Crystal structure of metallo DNA duplex containing consecutive Watson-Crick-like T-Hg(II)-T base pairs. *Angew. Chem. Int. Ed. Engl.*, **53**, 2385–2388.
39. Liu,H., Cai,C., Haruehanroengra,P., Yao,Q., Chen,Y., Yang,C., Luo,Q., Wu,B., Li,J., Ma,J *et al.*2017) Flexibility and stabilization of HgII-mediated C:T and T:T base pairs in DNA duplex. *Nucleic Acids Res.*, **45**, 2910–2918.
40. Dattagupta,N. and Crothers,D.M. (1981) Solution structural studies of the Ag(I)-DNA complex. *Nucleic Acids Res.*, **9**, 2971–2985.
41. Eichhorn,G.L., Butzow,J.J., Clark,P. and Tarien,E. (1967) Interaction of metal ions with polynucleotides and related compounds. X. Studies on the reaction of silver (I) with the nucleosides and polynucleotides, and the effect of silver(I) on the zinc(II) degradation of polynucleotides. *Biopolymers*, **5**, 283–296.
42. Johannsen,S., Megger,N., Böhme,D., Sigel,R.K.O. and Müller,J. (2010) Solution structure of a DNA double helix with consecutive metal-mediated base pairs. *Nat. Chem.*, **2**, 229.
43. Kumbhar,S., Johannsen,S., Sigel,R.K., Waller,M.P. and Muller,J. (2013) A QM/MM refinement of an experimental DNA structure with metal-mediated base pairs. *J. Inorg. Biochem.*, **127**, 203–210.
44. Liu,H., Shen,F., Haruehanroengra,P., Yao,Q., Cheng,Y., Chen,Y., Yang,C., Zhang,J., Wu,B., Luo,Q *et al.*2017) A DNA structure containing Ag(I) -mediated G:G and C:C base pairs. *Angew. Chem. Int. Ed. Engl.*, **56**, 9430–9434.
45. Kondo,J., Tada,Y., Dairaku,T., Hattori,Y., Saneyoshi,H., Ono,A. and Tanaka,Y. (2017) A metallo-DNA nanowire with uninterrupted one-dimensional silver array. *Nat. Chem.*, **9**, 956.
46. Guo,L., Hu,H., Sun,R. and Chen,G. (2009) Highly sensitive fluorescent sensor for mercury ion based on photoinduced charge transfer between fluorophore and pi-stacked T-Hg(II)-T base pairs. *Talanta*, **79**, 775–779.
47. Ito,T., Nikaido,G. and Nishimoto,S. (2007) Effects of metal binding to mismatched base pairs on DNA-mediated charge transfer. *J. Inorg. Biochem.*, **101**, 1090–1093.
48. Hiroyuki,I., Naomi,Y., Atsushi,A., Tomoko,F., Waka,N. and Shu,S. (2011) Electron mobility in a mercury-mediated duplex of triazole-linked DNA (TLDNA). *Chem. Lett.*, **40**, 318–319.
49. Schmidt,M.H. and Pearson,C.E. (2016) Disease-associated repeat instability and mismatch repair. *DNA Repair (Amst.)*, **38**, 117–126.
50. Mirkin,S.M. (2006) DNA structures, repeat expansions and human hereditary disorders. *Curr. Opin. Struct. Biol.*, **16**, 351–358.
51. Zhao,X.N. and Usdin,K. (2015) The repeat expansion diseases: the dark side of DNA repair. *DNA Repair (Amst.)*, **32**, 96–105.
52. Gatchel,J.R. and Zoghbi,H.Y. (2005) Diseases of unstable repeat expansion: mechanisms and common principles. *Nat. Rev. Genet.*, **6**, 743–755.
53. Chen,Y.W., Jhan,C.R., Neidle,S. and Hou,M.H. (2014) Structural basis for the identification of an i-motif tetraplex core with a parallel-duplex junction as a structural motif in CCG triplet repeats. *Angew. Chem. Int. Ed. Engl.*, **53**, 10682–10686.
54. Palumbo,S.L., Memmott,R.M., Uribe,D.J., Krotova-Khan,Y., Hurley,L.H. and Ebbinghaus,S.W. (2008) A novel G-quadruplex-forming GGA repeat region in the c-myb promoter is a critical regulator of promoter activity. *Nucleic Acids Res.*, **36**, 1755–1769.
55. Bzymek,M. and Lovett,S.T. (2001) Instability of repetitive DNA sequences: the role of replication in multiple mechanisms. *PNAS*, **98**, 8319–8325.
56. Charlesworth,B., Sniegowski,P. and Stephan,W. (1994) The evolutionary dynamics of repetitive DNA in eukaryotes. *Nature*, **371**, 215–220.
57. Kiliszek,A. and Rypniewski,W. (2014) Structural studies of CNG repeats. *Nucleic Acids Res.*, **42**, 8189–8199.
58. Polyzos,A.A. and McMurray,C.T. (2017) Close encounters: moving along bumps, breaks, and bubbles on expanded trinucleotide tracts. *DNA Repair (Amst.)*, **56**, 144–155.
59. Thirugnanasambandam,A., Karthik,S., Mandal,P.K. and Gautham,N. (2015) The novel double-folded structure of d(GCATGCATGC): a possible model for triplet-repeat sequences. *Acta Crystallogr. D, Biol. Crystallogr.*, **71**, 2119–2126.
60. Viladoms,J., Escaja,N., Frieden,M., Gomez-Pinto,I., Pedroso,E. and Gonzalez,C. (2009) Self-association of short DNA loops through minor groove C:G:G:C tetrads. *Nucleic Acids Res.*, **37**, 3264–3275.
61. Thorpe,J.H., Teixeira,S.C., Gale,B.C. and Cardin,C.J. (2003) Crystal structure of the complementary quadruplex formed by d(GCATGCT) at atomic resolution. *Nucleic Acids Res.*, **31**, 844–849.
62. Leonard,G.A., Zhang,S., Peterson,M.R., Harrop,S.J., Helliwell,J.R., Cruse,W.B., d'Estaintot,B.L., Kennard,O., Brown,T. and Hunter,W.N. (1995) Self-association of a DNA loop creates a quadruplex: crystal structure of d(GCATGCT) at 1.8 A resolution. *Structure*, **3**, 335–340.
63. Salisbury,S.A., Wilson,S.E., Powell,H.R., Kennard,O., Lubini,P., Sheldrick,G.M., Escaja,N., Alazzouzi,E., Grandas,A. and Pedroso,E. (1997) The bi-loop, a new general four-stranded DNA motif. *PNAS*, **94**, 5515–5518.
64. Huang,T.Y., Chang,C.K., Kao,Y.F., Chin,C.H., Ni,C.W., Hsu,H.Y., Hu,N.J., Hsieh,L.C., Chou,S.H., Lee,I.R *et al.*2017) Parity-dependent hairpin configurations of repetitive DNA sequence promote slippage associated with DNA expansion. *PNAS*, **114**, 9535–9540.
65. Lam,S.L., Wu,F., Yang,H. and Chi,L.M. (2011) The origin of genetic instability in CCTG repeats. *Nucleic Acids Res.*, **39**, 6260–6268.
66. Guo,P. and Lam,S.L. (2016) Minidumbbell: a new form of native DNA structure. *J. Am. Chem. Soc.*, **138**, 12534–12540.
67. Du,Y.H., Huang,J., Weng,X.C. and Zhou,X. (2010) Specific recognition of DNA by small molecules. *Curr. Med. Chem.*, **17**, 173–189.
68. Rhee,K.Y., Senear,D.F. and Hatfield,G.W. (1998) Activation of gene expression by a ligand-induced conformational change of a protein-DNA complex. *J. Biol. Chem.*, **273**, 11257–11266.
69. Zinkel,S.S. and Crothers,D.M. (1991) Catabolite activator protein-induced DNA bending in transcription initiation. *J. Mol. Biol.*, **219**, 201–215.
70. Rice,P.A., Yang,S., Mizuuchi,K. and Nash,H.A. (1996) Crystal structure of an IHF-DNA complex: a protein-induced DNA U-turn. *Cell*, **87**, 1295–1306.
71. Kim,J., Zwieb,C., Wu,C. and Adhya,S. (1989) Bending of DNA by gene-regulatory proteins: construction and use of a DNA bending vector. *Gene*, **85**, 15–23.

72. Ha,S.C., Lowenhaupt,K., Rich,A., Kim,Y.G. and Kim,K.K. (2005) Crystal structure of a junction between B-DNA and Z-DNA reveals two extruded bases. *Nature*, **437**, 1183–1186.
73. Almaqwashi,A.A., Paramanathan,T., Rouzina,I. and Williams,M.C. (2016) Mechanisms of small molecule-DNA interactions probed by single-molecule force spectroscopy. *Nucleic Acids Res.*, **44**, 3971–3988.
74. Chen,Y.W. and Hou,M.H. (2013) The binding of the Co(II) complex of dimeric chromomycin A3 to GC sites with flanking G:G mismatches. *J. Inorg. Biochem.*, **121**, 28–36.
75. Campbell,V.W., Davin,D., Thomas,S., Jones,D., Roesel,J., Tran-Patterson,R., Mayfield,C.A., Rodu,B., Miller,D.M. and Hiramoto,R.A. (1994) The G-C specific DNA binding drug, mithramycin, selectively inhibits transcription of the C-MYC and C-HA-RAS genes in regenerating liver. *Am. J. Med. Sci.*, **307**, 167–172.
76. Hsu,C.W., Chuang,S.M., Wu,W.L. and Hou,M.H. (2012) The crucial role of divalent metal ions in the DNA-acting efficacy and inhibition of the transcription of dimeric chromomycin A3. *PLoS One*, **7**, e43792.
77. Hou,M.H., Robinson,H., Gao,Y.G. and Wang,A.H. (2004) Crystal structure of the [Mg2+-(chromomycin A3)2]-d(TTGGCCAA)2 complex reveals GGCC binding specificity of the drug dimer chelated by a metal ion. *Nucleic Acids Res.*, **32**, 2214–2222.
78. Hou,C., Weidenbach,S., Cano,K.E., Wang,Z., Mitra,P., Ivanov,D.N., Rohr,J. and Tsodikov,O.V. (2016) Structures of mithramycin analogues bound to DNA and implications for targeting transcription factor FLI1. *Nucleic Acids Res.*, **44**, 8990–9004.
79. Waring,M.J. and Wakelin,L.P. (1974) Echinomycin: a bifunctional intercalating antibiotic. *Nature*, **252**, 653–657.
80. Pfoh,R., Cuesta-Seijo,J.A. and Sheldrick,G.M. (2009) Interaction of an echinomycin-DNA complex with manganese ions. *Acta Crystallogr. F, Struct. Biol. Crystall. Commun.*, **65**, 660–664.
81. Cuesta-Seijo,J.A., Weiss,M.S. and Sheldrick,G.M. (2006) Serendipitous SAD phasing of an echinomycin-(ACGTACGT)2 bisintercalation complex. *Acta Crystallogr. D, Biol. Crystallogr.*, **62**, 417–424.
82. Cuesta-Seijo,J.A. and Sheldrick,G.M. (2005) Structures of complexes between echinomycin and duplex DNA. *Acta Crystallogr. D, Biol. Crystallogr.*, **61**, 442–448.
83. Wu,P.C., Tzeng,S.L., Chang,C.K., Kao,Y.F., Waring,M.J. and Hou,M.H. (2018) Cooperative recognition of T:T mismatch by echinomycin causes structural distortions in DNA duplex. *Nucleic Acids Res.*, doi:10.1093/nar/gky345.
84. Jarjanazi,H., Li,H., Andrulis,I.L. and Ozcelik,H. (2006) Genome wide screening of CAG trinucleotide repeat lengths in breast cancer. *Dis. Markers*, **22**, 343–349.
85. Thion,M.S., Tezenas du Montcel,S., Golmard,J.L., Vacher,S., Barjhoux,L., Sornin,V., Cazeneuve,C., Bieche,I., Sinilnikova,O., Stoppa-Lyonnet,D *et al.*2016) CAG repeat size in Huntingtin alleles is associated with cancer prognosis. *Eur. J. Hum. Genet: EJHG*, **24**, 1310–1315.
86. Vos,J.G. and Kelly,J.M. (2006) Ruthenium polypyridyl chemistry; from basic research to applications and back again. *Dalton Trans*, 4869–4883.
87. Hall,J.P., Keane,P.M., Beer,H., Buchner,K., Winter,G., Sorensen,T.L., Cardin,D.J., Brazier,J.A. and Cardin,C.J. (2016) Delta chirality ruthenium 'light-switch' complexes can bind in the minor groove of DNA with five different binding modes. *Nucleic Acids Res.*, **44**, 9472–9482.
88. Hall,J.P., O'Sullivan,K., Naseer,A., Smith,J.A., Kelly,J.M. and Cardin,C.J. (2011) Structure determination of an intercalating ruthenium dipyridophenazine complex which kinks DNA by semiintercalation of a tetraazaphenanthrene ligand. *PNAS*, **108**, 17610–17614.
89. Hall,J.P., Sanchez-Weatherby,J., Alberti,C., Quimper,C.H., O'Sullivan,K., Brazier,J.A., Winter,G., Sorensen,T., Kelly,J.M., Cardin,D.J *et al.*2014) Controlled dehydration of a ruthenium complex-DNA crystal induces reversible DNA kinking. *J. Am. Chem. Soc.*, **136**, 17505–17512.
90. Hsu,C.F., Phillips,J.W., Trauger,J.W., Farkas,M.E., Belitsky,J.M., Heckel,A., Olenyuk,B.Z., Puckett,J.W., Wang,C.C. and Dervan,P.B. (2007) Completion of a Programmable DNA-Binding Small Molecule Library. *Tetrahedron*, **63**, 6146–6151.
91. Nanjunda,R. and Wilson,W.D. (2012) Binding to the DNA minor groove by heterocyclic dications: from AT-specific monomers to GC recognition with dimers. *Curr. Protoc. Nucleic Acid Chem.*, doi:10.1002/0471142700.nc0808s51.
92. Buchmueller,K.L., Staples,A.M., Uthe,P.B., Howard,C.M., Pacheco,K.A., Cox,K.K., Henry,J.A., Bailey,S.L., Horick,S.M., Nguyen,B *et al.*2005) Molecular recognition of DNA base pairs by the formamido/pyrrole and formamido/imidazole pairings in stacked polyamides. *Nucleic Acids Res.*, **33**, 912–921.
93. White,S., Baird,E.E. and Dervan,P.B. (1996) Effects of the A.T/T.A degeneracy of pyrrole–imidazole polyamide recognition in the minor groove of DNA. *Biochemistry*, **35**, 12532–12537.
94. Chenoweth,D.M. and Dervan,P.B. (2009) Allosteric modulation of DNA by small molecules. *PNAS*, **106**, 13175–13179.
95. Chenoweth,D.M. and Dervan,P.B. (2010) Structural basis for cyclic Py-Im polyamide allosteric inhibition of nuclear receptor binding. *J. Am. Chem. Soc.*, **132**, 14521–14529.
96. Nickols,N.G. and Dervan,P.B. (2007) Suppression of androgen receptor-mediated gene expression by a sequence-specific DNA-binding polyamide. *PNAS*, **104**, 10418–10423.
97. Diaz-Perez,S., Kane,N., Kurmis,A.A., Yang,F., Kummer,N.T., Dervan,P.B. and Nickols,N.G. (2018) Interference with DNA repair after ionizing radiation by a pyrrole-imidazole polyamide. *PLoS One*, **13**, e0196803.
98. Pett,L., Kiakos,K., Satam,V., Patil,P., Laughlin-Toth,S., Gregory,M., Bowerman,M., Olson,K., Savagian,M., Lee,M *et al.*2017) Modulation of topoisomerase IIalpha expression and chemosensitivity through targeted inhibition of NF-Y:DNA binding by a diamino p-anisyl-benzimidazole (Hx) polyamide. *Biochim. Biophys. Acta*, **1860**, 617–629.
99. Kageyama,Y., Sugiyama,H., Ayame,H., Iwai,A., Fujii,Y., Huang,L.E., Kizaka-Kondoh,S., Hiraoka,M. and Kihara,K. (2006) Suppression of VEGF transcription in renal cell carcinoma cells by pyrrole-imidazole hairpin polyamides targeting the hypoxia responsive element. *Acta Oncol*, **45**, 317–324.
100. Nickols,N.G., Jacobs,C.S., Farkas,M.E. and Dervan,P.B. (2007) Modulating hypoxia-inducible transcription by disrupting the HIF-1-DNA interface. *ACS Chem. Biol.*, **2**, 561–571.
101. Padroni,G., Parkinson,J.A., Fox,K.R. and Burley,G.A. (2018) Structural basis of DNA duplex distortion induced by thiazole-containing hairpin polyamides. *Nucleic Acids Res.*, **46**, 42–53.
102. Nakatani,K., Hagihara,S., Goto,Y., Kobori,A., Hagihara,M., Hayashi,G., Kyo,M., Nomura,M., Mishima,M. and Kojima,C. (2005) Small-molecule ligand induces nucleotide flipping in (CAG)n trinucleotide repeats. *Nat. Chem. Biol.*, **1**, 39–43.
103. Tseng,W.H., Chang,C.K., Wu,P.C., Hu,N.J., Lee,G.H., Tzeng,C.C., Neidle,S. and Hou,M.H. (2017) Induced-Fit Recognition of CCG Trinucleotide Repeats by a Nickel-Chromomycin Complex Resulting in Large-Scale DNA Deformation. *Angew. Chem. Int. Ed. Engl.*, **56**, 8761–8765.
104. Pierre,V.C., Kaiser,J.T. and Barton,J.K. (2007) Insights into finding a mismatch through the structure of a mispaired DNA bound by a rhodium intercalator. *PNAS*, **104**, 429–434.
105. Cordier,C., Pierre,V.C. and Barton,J.K. (2007) Insertion of a bulky rhodium complex into a DNA cytosine-cytosine mismatch: an NMR solution study. *J. Am. Chem. Soc.*, **129**, 12287–12295.
106. Zeglis,B.M., Pierre,V.C., Kaiser,J.T. and Barton,J.K. (2009) A bulky rhodium complex bound to an adenosine-adenosine DNA mismatch: general architecture of the metalloinsertion binding mode. *Biochemistry*, **48**, 4247–4253.
107. Song,H., Kaiser,J.T. and Barton,J.K. (2012) Crystal structure of Delta-[Ru(bpy)(2)dppz](2)(+) bound to mismatched DNA reveals side-by-side metalloinsertion and intercalation. *Nat. Chem.*, **4**, 615–620.
108. Lo,Y.S., Tseng,W.H., Chuang,C.Y. and Hou,M.H. (2013) The structural basis of actinomycin D-binding induces nucleotide flipping out, a sharp bend and a left-handed twist in CGG triplet repeats. *Nucleic Acids Res.*, **41**, 4284–4294.
109. Woo,C.M., Li,Z., Paulson,E.K. and Herzon,S.B. (2016) Structural basis for DNA cleavage by the potent antiproliferative agent (-)-lomaiviticin A. *PNAS*, **113**, 2851–2856.

110. Jourdan,M., Granzhan,A., Guillot,R., Dumy,P. and Teulade-Fichou,M.P. (2012) Double threading through DNA: NMR structural study of a bis-naphthalene macrocycle bound to a thymine-thymine mismatch. *Nucleic Acids Res.*, **40**, 5115–5128.
111. Whelan,D.R., Hiscox,T.J., Rood,J.I., Bambery,K.R., McNaughton,D. and Wood,B.R. (2014) Detection of an en masse and reversible B- to A-DNA conformational transition in prokaryotes in response to desiccation. *J. R. Soc. Interface*, **11**, 20140454.
112. Wittig,B., Dorbic,T. and Rich,A. (1991) Transcription is associated with Z-DNA formation in metabolically active permeabilized mammalian cell nuclei. *PNAS*, **88**, 2259–2263.
113. Rich,A. and Zhang,S. (2003) Timeline: Z-DNA: the long road to biological function. *Nat. Rev. Genet.*, **4**, 566–572.
114. Mulholland,N., Xu,Y., Sugiyama,H. and Zhao,K. (2012) SWI/SNF-mediated chromatin remodeling induces Z-DNA formation on a nucleosome. *Cell Biosci.*, **2**, 3.
115. Hou,X.M., Fu,Y.B., Wu,W.Q., Wang,L., Teng,F.Y., Xie,P., Wang,P.Y. and Xi,X.G. (2017) Involvement of G-triplex and G-hairpin in the multi-pathway folding of human telomeric G-quadruplex. *Nucleic Acids Res.*, **45**, 11401–11412.
116. Juen,M.A., Wunderlich,C.H., Nussbaumer,F., Tollinger,M., Kontaxis,G., Konrat,R., Hansen,D.F. and Kreutz,C. (2016) Excited states of nucleic acids probed by proton relaxation dispersion NMR spectroscopy. *Angew. Chem. Int. Ed. Engl.*, **55**, 12008–12012.
117. Dans,P.D., Ivani,I., Hospital,A., Portella,G., Gonzalez,C. and Orozco,M. (2017) How accurate are accurate force-fields for B-DNA? *Nucleic Acids Res.*, **45**, 4217–4230.
118. Yang,D., Tan,Z., Mi,Y. and Wei,B. (2017) DNA nanostructures constructed with multi-stranded motifs. *Nucleic Acids Res.*, **45**, 3606–3611.
119. Furst,A., Hill,M.G. and Barton,J.K. (2014) Electrocatalysis in DNA sensors. *Polyhedron*, **84**, 150–159.
120. Service,R.F. (2011) DNA nanotechnology. Next step: DNA robots? *Science*, **332**, 1142.