Briefings in Functional Genomics, 18(3), 2019, 192–204

doi: 10.1093/bfgp/ely026
Advance Access Publication Date: 8 August 2018
Review paper

Dynamics of physiologically relevant noncanonical DNA structures: an overview from experimental and theoretical studies

Debostuti Ghoshdastidar and Manju Bansal

Corresponding author: Manju Bansal, Molecular Biophysics Unit, Indian Institute of Science, Bangalore 560 012, India. Tel: 91-80-22932534; Fax: 91-80-23600535; E-mail: mb@iisc.ac.in

Abstract

DNA is a complex molecule with phenomenal inherent plasticity and the ability to form different hydrogen bonding patterns of varying stabilities. These properties enable DNA to attain a variety of structural and conformational polymorphic forms. Structurally, DNA can exist in single-stranded form or as higher-order structures, which include the canonical double helix as well as the noncanonical duplex, triplex and quadruplex species. Each of these structural forms in turn encompasses an ensemble of dynamically heterogeneous conformers depending on the sequence composition and environmental context. *In vivo*, the widely populated canonical B-DNA attains these noncanonical polymorphs during important cellular processes. While several investigations have focused on the structure of these noncanonical DNA, studying their dynamics has remained nontrivial. Here, we outline findings from some recent advanced experimental and molecular simulation techniques that have significantly contributed toward understanding the complex dynamics of physiologically relevant noncanonical forms of DNA.

Key words: dynamics of noncanonical DNA; NMR spectroscopy; molecular dynamics simulations; DNA structural polymorphism; DNA conformational polymorphism; higher-order DNA structure

Introduction

Detailed elucidation of the nature, composition and structure of nucleic acids occurred concurrently in the mid twentieth century by independently working groups of physiologists, chemists and biophysicists. Nucleic acids were found to be of two types, DNA and RNA, differing primarily in their chemical nature but also in their cellular localization. The 1st correct three-dimensional structure of nucleic acids was the double-helical model of DNA proposed by James Watson and Francis Crick in 1953 (Figure 1A) [1]. Over 6 decades and 40 000 publications later, the double helix has stood the test of time. In addition, a huge body of research has revealed the extraordinary structural and conformational plasticity inherent in the conventional Watson–Crick (WC) model of DNA [2–7]. The apparently simple, four-letter coded biopolymer is now known to exist in a plethora of different conformational variants, with varying structural features, and forms several higher-order structures. This versatility of DNA structure and conformation underlies the specificity of transcription, the fidelity of replication and regulation of gene activity. In the context of applications, nucleic acids form the central theme of areas as diverse as molecular biology, pharmaceutical technology, nanotechnology and digital information storage [8–13]. Concurrently, RNA structure and biology have evolved into an active field owing to the even larger conformational and functional diversity exhibited by this class of nucleic acids. However, RNA conformation, dynamics and

Debostuti Ghoshdastidar obtained her doctoral degree from the Indian Institute of Technology Madras, India. As a postdoctoral fellow in Prof. Manju Bansal's laboratory she is studying the dynamics of DNA-transcription factor binding.
Manju Bansal was in the faculty of IISc from 1982–2016 and is a J.C. Bose National Fellow. Her research relates protein/DNA sequences to structure and function with development of computational algorithms and tools.

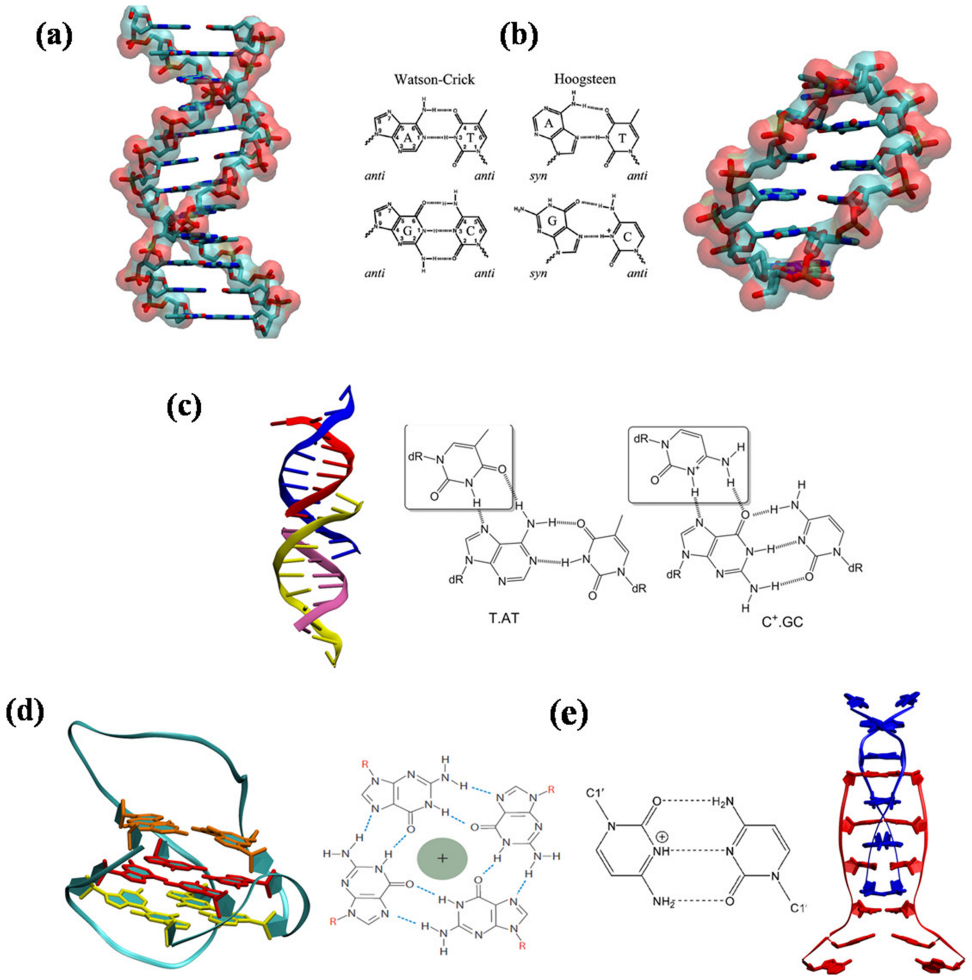

Figure 1. Canonical and noncanonical forms of DNA. A DNA double helix with sequence poly(AT) can form both (**A**) WC as well as (**B**) HG H-bonds. The corresponding duplex structures of WC dodecamer (created using NAB utility [127]) and HG hexamer (PDBid:2qs6) are shown. DNA can form various higher-order structures like (**C**) triplex (PDBid:1d3r), (**D**) G-quadruplex (PDBid:2f8u) and (**E**) i-motif (PDBid:1ybn).

function have been dealt elsewhere in this volume and, hence, will not be discussed here.

The canonical WC double helix can exist in several structural and conformational forms based on the sequence, solvation environment and interaction partners, with several of them playing important functional roles *in vivo* [3, 4]. The structural polymorphs of double-helical DNA emanate from the conformational flexibility of the phosphodiester backbone, including the puckering of the deoxyribose sugar and rotations about the glycosyl bond, as well as the variable hydrogen bonding abilities of the nucleobases [3]. Despite large differences in their structural parameters, all the canonical variants are duplex with two strands held together by inter-strand WC hydrogen bonding between purine and pyrimidine bases, forming A-T and G-C pairs (Figure 1A). While studies on the canonical double helix dominated the period following its elucidation, eventually several other hydrogen-bonding arrangements between base pairs were observed. In principle, at least 28 base-pairing patterns are possible, with at least 2 H-bonds between the paired bases in their 'keto' form, though they can exist in alternate tautomeric forms ('keto' or 'enol') [14, 15]. The most significant alternative pattern of pairing A-T and G-C bases was first observed by Karst Hoogsteen in 1959 and named after the discoverer as Hoogsteen (HG) bps [16]. In a duplex, HG bps are formed by 180-degree rotation of the purine base around the glycosidic bond to form a syn conformation compared to the 'anti'-conformation observed in the WC bp (Figure 1B). This results in the HG double helix having features distinct from the classical B-form, like a narrower minor groove, shorter C1′-C1′ distance across a base pair, and higher number of hydrogen bonding sites at the major groove edge. Consequently, HG bps are often accommodated well near structurally stressed regions of DNA that result from DNA lesions, protein or antibiotic binding, DNA supercoiling or near nicks and loops on the WC double helix [17–20 and references therein].

While duplex DNA is the most widespread and physiologically relevant *in vivo* form, DNA exists in non-duplex forms during cellular events. For example, the single-stranded DNA (ssDNA) is generated as an intermediate during several cellular processes like replication, transcription, recombination and repair [21, 22]. *In vivo*, ssDNA almost always exists in complex with protein, with the exception of ssDNA viruses that contain genomes composed of kilo bp-length ssDNA. Accumulation of naked ssDNA has therefore been implicated in DNA lesions and failure of DNA repair mechanism. ssDNA can also form hairpin structures that are involved in replication, transcription and recombination in both prokaryotic and eukaryotic systems. ssDNA hairpin formation has been associated with triple-helical

repeat disorders like Fragile X syndrome and other neurological disorders [23, 24]. Thus, investigation of conformations and interactions of ssDNA has been an active area of research over several decades.

The conformational plasticity of DNA and its ability to form stable hydrogen bonds (both WC and HG) enables the formation of higher-order DNA structures. Alternate double-stranded structures, like cruciform DNA, are formed by inverted repeat sequences and stabilized by negative supercoiling [25]. Cruciform DNAs play regulatory role in vital biological processes including replication, regulation of gene expression, nucleosome structure and recombination and have therefore been implicated in lethal diseases like cancer. Triplex or triple-helical DNA is formed when homopurine–homopyrimidine oligonucleotide strands forming a WC double helix engage a 3rd homopyrimidine strand in its major groove via HG or reverse HG base pairing (Figure 1C) [6]. The sequence-specific nature of triplex formation has shown great potential in oligonucleotide-mediated site directed recombination, mutagen delivery and most importantly in antigene therapy for targeting diseased genes and chromosomes [26–28]. Genomic studies have predicted the occurrence of one triplex forming site every 1366 bases in the human genome, indicating that oligonucleotide-mediated triplex formation can be carried out at a high resolution [29]. However, none of these studies have provided evidence for the presence of triplex DNA inside the living cell. On the other hand, quadruplex forms of DNA have been observed *in vivo* and characterized widely for their biological role. *In vivo* a G-rich DNA strand can fold over itself to form the characteristic G-quadruplex (GQ), while the complementary C-rich strand forms the i-motif (Figures 1D and E) [2, 30–36]. Computational analyses of genome sequences have predicted the presence of $>$ 3,75,000 quadruplex-forming motifs in human and $>$ 1400 in the yeast genome [35 and references therein]. These motifs are found to cluster in key gene regulatory regions like untranslated regions (UTRs), telomeres, oncogene promoters, immunoglobulin switch regions, recombination hot spots, centromeres and in fragile X-syndrome, where their occurrence is highly conserved [37, 38]. Hence, regulation of quadruplex formation has been implicated in cancer and other diseases. Moreover, stabilization of G-quartet and i-motifs at telomeres has also shown potential in anti-aging therapy [39].

The brief account presented above illustrates that noncanonical forms of DNA play a significant role in vital cellular processes. However, they are present under specific physiological conditions inside the living cell. As a result, detection of these unusual DNA structures and studying their dynamics are not trivial tasks. Nonetheless, the intimate correlation between conformation and structure is central to unraveling the functional roles of biomolecules. Therefore, a considerable body of research has focused on investigating unconventional nucleic acid structures using primarily spectroscopic and molecular dynamics (MD) simulation techniques. In the rest of the review we have attempted to summarize key findings on the detection and dynamics of noncanonical DNA forms, which are physiologically relevant.

Conformational Polymorphism in ssDNA

Dynamical conformers of homopolymeric ssDNA

The ssDNA, unlike its duplex counterpart, is not constrained by inter-strand hydrogen bonding and thereby explores a large conformational space owing to the flexibility of the polynucleotide backbone. The conformations attained by ssDNA are further dependent on the oligonucleotide sequence and its interaction with the solvent and counterion environment [40, 41]. Single-molecule force extension spectroscopic studies have shown that persistence length of DNA, which is inversely correlated with DNA flexibility, is several order of magnitude smaller for ssDNA compared to dsDNA [42]. Since investigation of such diverse conformational space is difficult with experimental techniques, computational statistical simulations like MD techniques have been employed. Indeed, MD simulations have revealed rapid dynamics of ssDNA, which progressively increases with the chain length of the oligonucleotide. The sugar phosphate backbone dihedrals of ssDNA also span broad distributions indicating the flexibility of the ssDNA compared to its duplex counterpart [40]. As shown in Figure 2A, MD simulations on a single strand of the Dickerson–Drew dodecamer showed the molecule undergoing large conformational changes, transforming from an ordered base-stacked form to a globular coil structure, in the nanosecond time scale [43]. Counterions stabilize the globular conformation by neutralizing intra-strand repulsion between phosphate backbones thereby favoring ssDNA folding [44]. The conformational transition is mediated by formation of base-stacking interactions between non-consecutive bases (i with $i + 2$ or $i + 3$). The propensity to form such non-sequential base stacking was found to be dependent on the sequence and length of the oligonucleotide. Thus, polyA oligonucleotides with increasing chain length preferably form base stacks and thereby exist in a globular coil conformation [40].

Dynamical conformers of heterogeneous ssDNA

While homopolymeric ssDNA transits into coiled conformations, heterogeneous ssDNA attains more well-defined secondary structures like hairpins, loops and pseudoknots. Among these, the structure, dynamics and biological role of the hairpin are well established. As shown in Figure 2B, the DNA hairpin comprises of two parts, a WC base-paired stem region and a loop region comprising of unpaired or non-WC base-paired nucleotides [45]. The stability of the hairpin structure is closely related with the oligonucleotide sequence, stem length and loop radius. Figure 2B presents a schematic representation of the transition pathway of an ssDNA (5′-GGATAA(T4)TTATCC-3′) from a random coil to a fully formed hairpin, obtained from laser temperature jump experiments in conjunction with fluorescence spectroscopic studies. As shown in the figure, following an initial nucleation step, the random coiled ssDNA gives rise to an ensemble of looped conformations. However, only the correctly folded loops permit folding into the native hairpin structure following a ‘zipping’ process to form the hairpin stem. Misfolded loops can lead to a kinetic trap and significantly delay the folding process, also resulting in anomalous folding kinetics frequently observed in studies of DNA hairpin formation. Thus, even short hairpins $<$10 bps in the stem unfold on time scales of millisecond or longer [46, 47].

The ability of ssDNA to exist as different three-dimensional conformers depending on its sequence forms the underlying basis of the single-stranded conformational polymorphism (SSCP) analysis for detection of gene mutations. A slight alteration in the sequence (few point mutations) causes mutated DNA strands to attain a different predominant semi-stable secondary structural form or a structure similar to the wild-type ssDNA with subtle variations in the conformation. Through a set of elegant experiments, Nishigaki *et al.* and others showed that this conformational polymorphism among different ssDNA

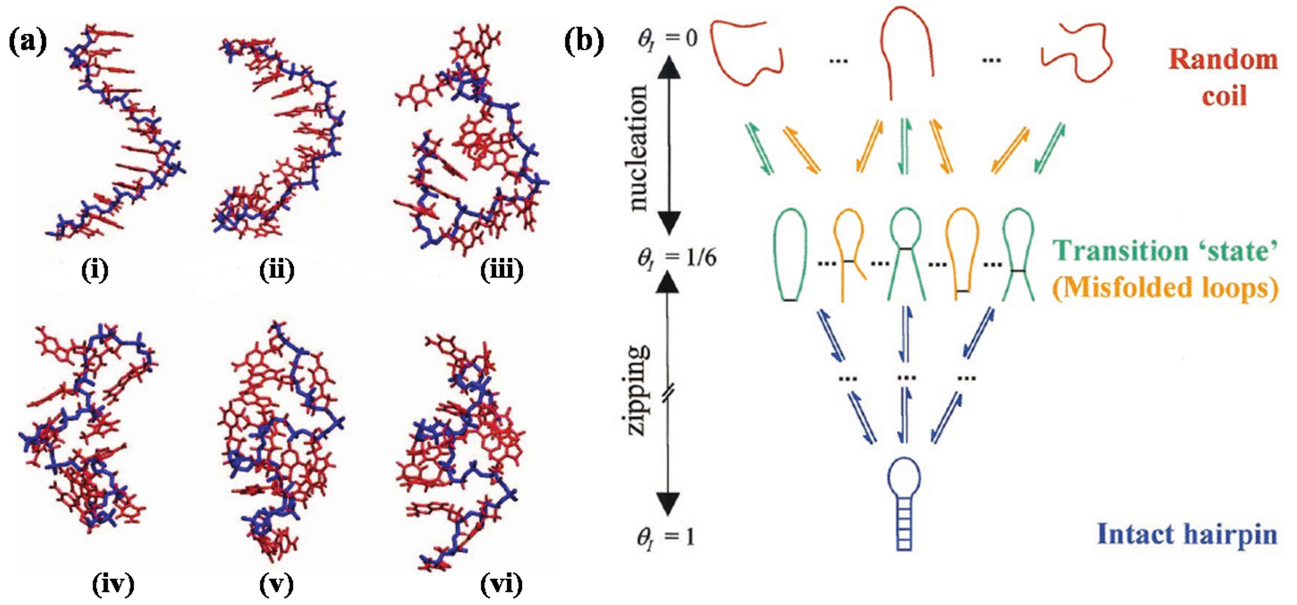

Figure 2. Conformational polymorphism exhibited by ssDNA. (**A**) Homopolymeric ssDNA traverses diverse conformational states from a fully extended form (i) to a collapsed coil during the course of MD simulation. Snapshots are shown for time points 10 (ii), 25 (iii), 50 (iv), 75 (v) and 100 ns (vi). (Reprinted from reference [39] with permission of AIP publishing) (**B**) A schematic representation of the transition pathway of heterogeneous ssDNA from a random coil to a fully formed hairpin shows that the DNA strand traverses through an ensemble of intermediate misfolded microstates, making the folding pathway more complex [45; Copyright (2001) National Academy of Sciences].

sequences results in significantly different electrophoretic migration and differential sensitivities to enzymatic cleavage [48–51]. Currently, over 2000 genes are known to carry at least one pathogenic mutation, and the number is steadily rising [52]. SSCP analyses, in conjunction with DNA sequencing, now form one of the most efficient tools for large-scale genetic diagnosis for identifying unknown pathogenic mutations.

Detection and Dynamics of Hoogsteen Base Pairs in Canonical Duplex DNA

Identification and characterization of the structure and biological role of the HG bp have an interesting history. The 1st crystal and solution Nuclear Magnetic Resonance (NMR) structures containing HG bps were obtained only for those DNA/ligand complexes in which the DNA duplex was either highly distorted or damaged [17 and references therein]. The simultaneous lack of evidence of HG bps in solution NMR structures of free DNA suggested two possibilities—either HG bps are artifacts that are stabilized by crystal packing effects or they are high energy conformations stabilized by severe distortions in DNA. These doubts were eventually put to rest with the identification of an HG bp embedded in the crystal structure of an undistorted B-DNA complexed with the homeodomain protein MATα2 (Figure 3) [20]. Subsequent NMR solution studies and MD simulations of uncomplexed DNA revealed that HG bps are often confined to highly flexible dinucleotide steps (e.g. CA/TG or TA/TA) and thereby possess the ability to rotate freely around the glycosidic bond to flip to the characteristic syn conformation, even in the absence of a ligand [17].

Recent advances in solution NMR spectroscopy provided a leap ahead in detecting transiently stable conformers in nucleic acids, often referred to as excited state (ES), relative to the canonical ground state (GS) WC DNA duplex [53]. Among these, developments in rotating frame (R1ρ) carbon relaxation dispersion (RD) NMR technique have been particularly useful in studying microsecond–millisecond dynamics under physiological conditions [54]. Investigations carried out by Al-Hashimi and co-workers using ^{13}C-R1ρ NMR RD revealed that HG bps form transiently across several dinucleotide steps, with populations ranging between 0.14 and 0.49% and lifetimes between 0.3 and 2.5 ms at pH 6.0 [55]. However, the relative thermodynamic stabilities of the HG bps varied significantly (from 2.1 to 2.8 kcal mol^{-1}) depending on the positional and

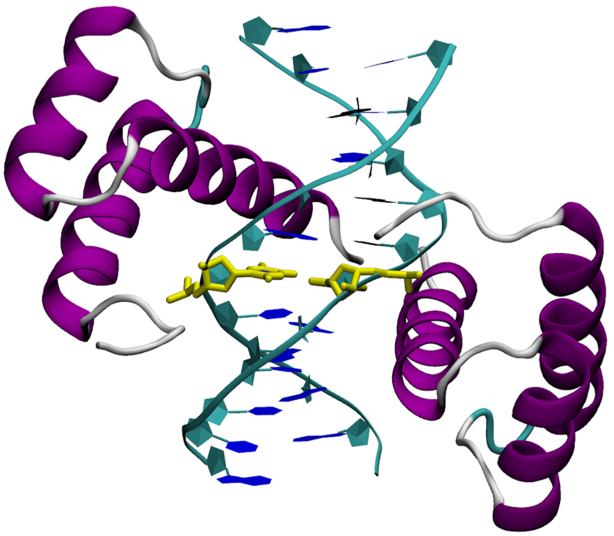

Figure 3. Crystal structure (PDBid:1k61) of homeodomain protein Matα2 with dsDNA shows that a AT HG bp (yellow) can be well accommodated within the canonical WC base-paired B-form without causing significant distortions in the global DNA structure [20].

sequence context. Moreover, the order of the stabilities of HG bps occurring at different dinucleotide steps was exactly the reverse of the order of stability of WC bps, indicating that less stable WC dinucleotide steps are more likely to harbor HG bps [55]. van't Hoff analysis of a temperature dependent RD data of GS (WC) <—> ES (HG) transition yielded activation-free energies of 16 kcal mol^{-1}, and a favorable enthalpy of 12– 26 kcal mol^{-1}. This indicates that the transition involves disruption of hydrogen bonds and base-stacking interactions, but this loss in enthalpy is restored once the HG bp is formed. The unfavorable entropy of transition, however, indicated that the ES (HG) could be more rigid than the GS (WC) conformation. MD simulation studies on duplex DNA containing HG bps confirmed high rigidity of HG duplexes, suggesting that the WC <—> HG transition is possibly achieved *in vivo* in the presence of an external factor that restricts the duplex flexibility, thereby compensating for the entropic preference for the B-form [56, 57].

Atomistic details of the WC–HG transition dynamics were explored by advanced MD techniques, since the large free energy barrier involved in the transition cannot be overcome using classical MD. In one of these investigations, a sample and select approach was followed in conjunction with MD simulations and NMR residual dipolar coupling (RDC) measurements to generate molecular snapshots of the transited DNA. Briefly, measured RDCs were used to guide selection of conformers from a pool of MD simulation generated snapshots such that the final ensemble of conformers encompasses the entire range of measured RDC data. Analyses of these structures revealed a 'kink-and-melt' dynamics of DNA transition. Formation of the HG bp was found to involve kinking of the duplex towards the major groove and a partial melting of the surrounding WC bps at the 3′-end [58]. The WC–HG transition dynamics were also explored by the conjugate peak refinement (CPR) and umbrella sampling methods [59, 60]. The CPR algorithm requires an initial guess of the minimum energy path connecting the initial and final states. In each CPR cycle, the path is modified by improving, removing or inserting one path-point so that the new path avoids the maximum energy peak. Repeating this heuristic procedure yields a final path that follows the valley of the energy surface by avoiding the high-energy points. The umbrella sampling technique overcomes activation energy barriers by applying a biasing potential along one or more reaction coordinates from the initial to the final state. The potential of mean force profile, i.e. the free energy obtained as a function of the reaction coordinate, can then be interpreted to obtain the transition pathway. In agreement with NMR studies, both simulation methods yielded transition pathways that involved disruption of WC base pairing and purine flip with low base pair opening angle towards the major groove, followed by formation of new hydrogen bonds to stabilize the HG base pair. It is important to note here that the widespread occurrence of HG bps in the genome had probably been overlooked by X-ray crystallographers due to the similar electron densities presented by HG and WC bps in DNA crystals, causing all base pairs to be modeled as WC base pairs by default. A recent comprehensive survey of the Protein Data Bank (PDB), using stringent geometric criteria to screen HG base pairs or intermediates of HG/WC bps in crystal structures, has found a total of 178 crystal structures containing HG and HG-like base pairs (0.3% of all crystal structures in PDB as of 2003) [61]. In agreement with experimental and MD simulation results, a large fraction of these base pairs are observed in duplex DNA complexed with protein or ligands [55–57]. Surprisingly, however, none of these base pairs are directly in contact with the bound ligand. Moreover, HG base pairs observed in naked DNA are found preferably in AT tracts.

The role of HG bps in vital cellular processes is now quite evident. HG bps in damaged DNA are specifically recognized by DNA repair enzymes, like MutS and DNA polymerases X and Y, enabling replication through DNA lesions [62]. A role of HG bps in pairing meiotic chromosomes during recombination has also been suggested [19]. HG bps present around the binding site in protein–DNA complexes appear to prevent steric clashes that could have hindered protein–DNA interactions in presence of WC bases. Finally, several noncanonical higher-order DNA structures like triple helical and quadruplex DNA, with extensive biological and therapeutic significance, are stabilized by the HG base pairing [17].

Detection and Dynamics of Quadruplex DNA Structures

Dynamics of DNA G-quadruplex

The fundamental building block of the GQ is a G-tetrad or a (G)-quartet, where one guanosine from each of the four strands is held together by four H-bonds between its WC face and the HG major groove face of a neighboring guanosine (Figure 1D) [6 and references therein]. A minimum of two tetrads must stack upon each other, stabilized by a centrally coordinated monovalent cation, to form a GQ. The sequence and length of the GQ, orientation of the guanosine nucleosides about the glycosidic bond ('syn' or 'anti'), topology of the extrahelical loops (parallel, diagonal or lateral), location of the coordinating cation, strand polarity (parallel or antiparallel) and consequently the structural parameters (groove width) can vary giving rise to a large number of distinct GQ structures [5, 63, 64]. Owing to such vast structural diversity for the same sequence, studying the conformational dynamics of GQ has proved challenging using standard biophysical experiments or classical unbiased atomistic simulations [65–68].

Very recently, however, there has been a surge in studies combining biophysical experiments with sophisticated statistical thermodynamic approaches to extract the thermodynamically and kinetically relevant pathway of GQ formation [69–74]. In addition, a significant number of studies have employed enhanced sampling MD techniques to identify the relevant intermediates in the folding pathway [65, 75–79]. The long time scale (several hours) of GQ folding and the long-lived stable intermediates observed in all the above studies suggest the kinetic partitioning (KP) mechanism of GQ formation [65]. The KP landscape is rugged, composed of several deep minima that thermodynamically resemble the native state closely and yet are separated from it by large free energy barriers. As a result, the folding mechanism involves several single-molecule misfolding–unfolding events with only a fraction of molecules attaining the native structure. Thermodynamic analyses have revealed that the complexity of the folding pathway, i.e. the number of folding intermediates, increases with increasing length of the G-motif [69]. Real-time NMR, florescence resonance energy transfer (FRET), electrospray ionisation mass spectrometry (ESI-MS) and circular dichroism (CD) spectroscopic studies have also confirmed the presence of multiple hybrid states in the folding pathway (Figure 4) [70–74]. As shown in Figure 4A, real-time NMR studies revealed that first a kinetically viable conformer is formed that eventually passes through an ensemble of unfolded states to attain the thermodynamically stable conformer [70]. Figure 4B presents CD spectra obtained during the course of folding induced by salt (KCl) titration in different G-forming motifs [71]. Each motif passes from an unfolded form in the absence of

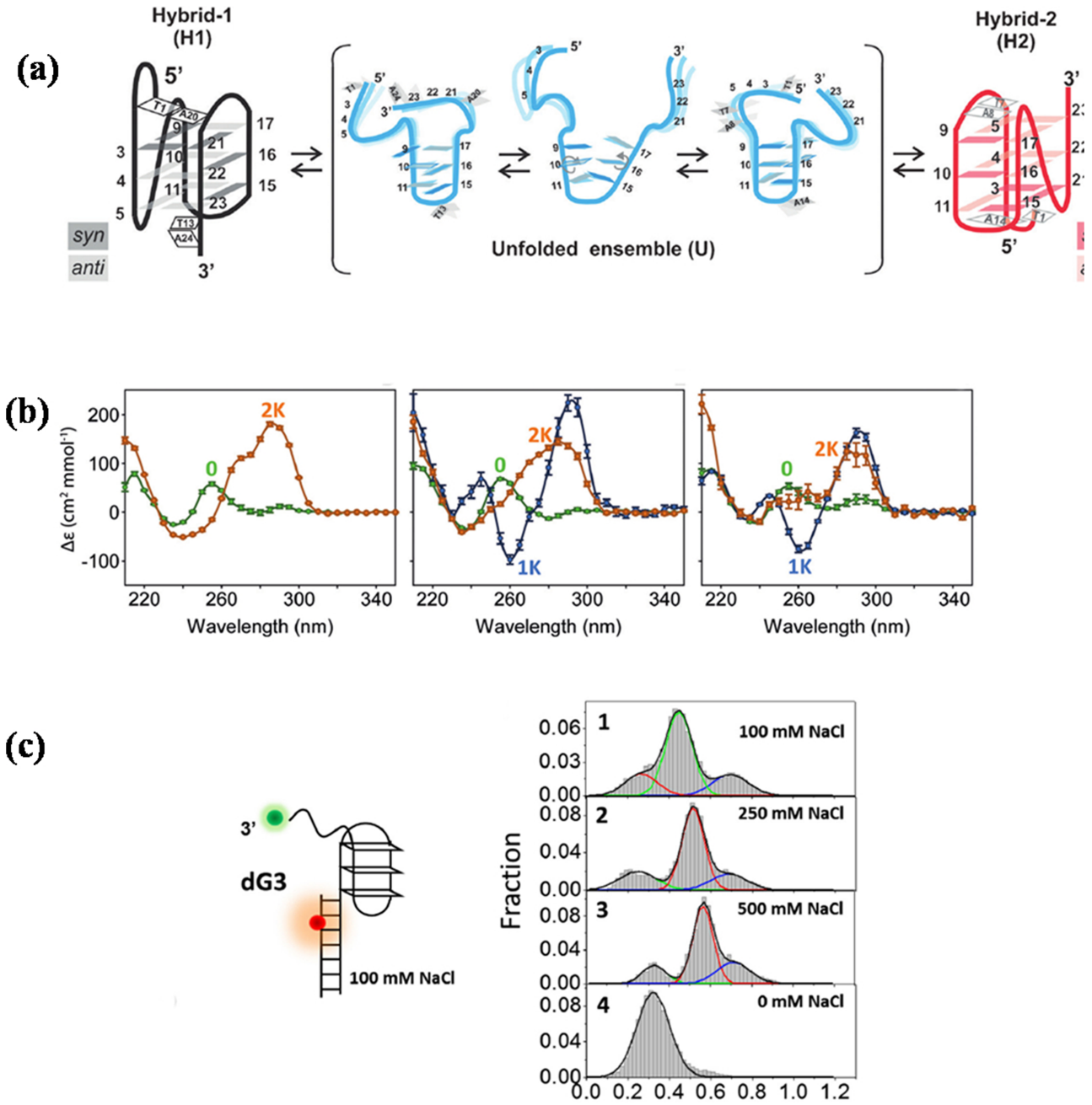

Figure 4. Experimental studies elucidate the complex folding pathway of GQ comprising of multiple folding intermediates. (**A**) Time-resolved (TR) NMR spectroscopic analyses of human telomeric G-motif reveal that a kinetically favored hybrid GQ conformer (H1) forms rapidly during the folding event and subsequently goes through a series of unfolding events to attain the final thermodynamically stable GQ form (H2).(Reproduced from [70] with permission from John Wiley & Sons) (**B**) CD spectra of K^+-induced folding of three different telomeric G-motifs show the presence of at least three conformers: single-stranded (0 K, green) in the absence of salt, antiparallel GQ (1 K, blue) at intermediate salt concentration and the final hybrid GQ (2 K, orange) at high salt concentration. The final conformation differs based on the sequence of the G-motif (compare the orange spectra) [71]. (**C**) FRET histograms show a single peak in absence of salt (0 mM NaCl), but multiple peaks in presence of salt indicating that the GQ formed are actually an ensemble of conformers and not a single conformer [73].

KCl (0 K, green graph) to the fully folded form at high KCl concentration (2 K, orange graph) through an intermediate antiparallel quadruplex conformer at intermediate KCl concentrations (1 K, blue graph). Interestingly, the diverse CD spectra obtained for the fully folded form of different G-motifs re-emphasize the structural heterogeneity exhibited by GQs. Presence of multiple conformers in the folding pathway was also confirmed by FRET studies on a human telomeric sequence (Figure 4C) [73]. The FRET histogram comprising of a single peak in the absence of salt, signifying the unfolded state, displayed multiple peaks at higher salt concentrations, indicating the presence of different polymorphic forms in the ion-induced GQ folding pathway. While all these experimental studies unequivocally established the presence of intermediates between the unfolded ssDNA state and the fully formed G-quartet, the structural characterization of these intermediates has remained elusive.

Computer simulations have achieved greater success in this regard. Three approaches have been primarily adopted by MD

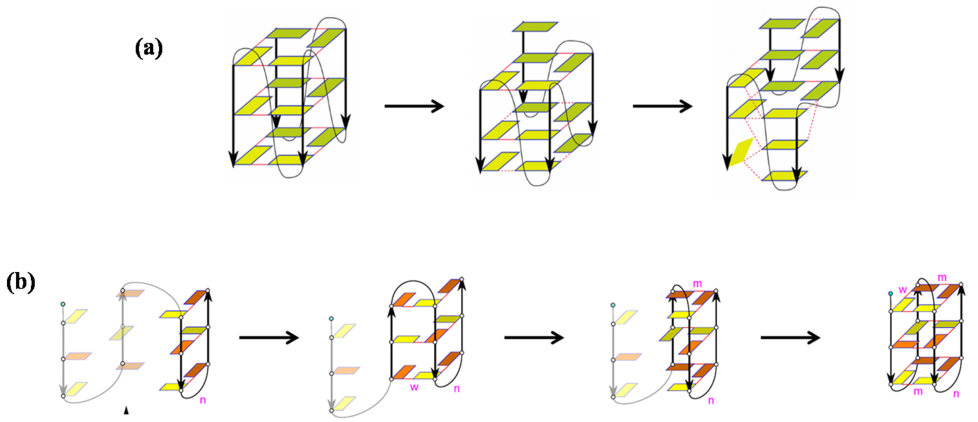

Figure 5. Computer simulations reveal structures of the possible intermediates in the GQ folding pathway. (**A**) The unfolding of a fully formed GQ (left panel) occurs through sequential single-strand slippage (middle and right panel). (**B**) The folding pathway comprises of hairpin (left panel) as early-folding intermediates and G-triplexes (middle panels) as late-folding intermediates of the final fully folded GQ (right panel) (Reprinted from reference [65] with permission from Elsevier).

studies to reach meaningful conclusions in characterizing the structural intermediates of GQ folding within the limits of time scale and computational resources. A brief overview of each method and its observed results are presented here. For a detailed review on the methodology and application of these techniques in investigating GQ-dynamics the reader is referred elsewhere [65]. The 1st approach involves the simulation of individual intermediates that have been speculated to be involved in GQ-folding from experiments [76]. By analyzing the structural dynamics and energetics of these intermediates, useful information is gained regarding their thermodynamic fitness in the folding pathway. Structures that are stable in the nanosecond time scale in explicit solvent simulations are regarded as potential intermediates while those that are unstable are disregarded. The 2nd method employed is standard MD simulations carried out under varying salt conditions [79]. Fully folded GQs are simulated in explicit solvent in the absence of ions (no-salt condition). Since GQs are unstable in the absence of ions, unfolding ensues, and its mechanism can thereby be studied. The unfolding pathway of GQ revealed by these two methods is in good agreement and has been summarized in Figure 5A. The GQ first undergoes strand slippage, giving rise to a G-triplex, which is followed by slippage of two other strands giving rise to a conformer with a single tetrad and three slipped strands. This unfolded form was then subjected to simulations under excess salt conditions to study the folding dynamics. The molecule embarked on a series of folding/misfolding events, during which the strands took up the ions and underwent back-slippage to finally reach close to the fully folded GQ state after several 100 nanoseconds of delay. The 3rd method makes use of enhanced sampling MD techniques like Replica Exchange MD and metadynamics [75, 77]. The GQ-folding pathway obtained for a 24-nt human telomeric DNA sequence $TTGGG(TTAGGG)_3$ using these techniques is summarized in Figure 5B. The single-stranded G-motif was found to pass through hairpin and triplex intermediates to form the final quadruplex structure.

Dynamics of DNA I-motif

Naturally occurring C-rich strands from human telomeric repeats, centromeric repeats, the human insulin minisatellite, the fragile X-repeat, oncogene promoters, etc. have been shown to fold into i-motif structures *in vitro* [2, 32, 80]. The peculiarity of the i-motif structure is the hemiprotonated C-C^+ mismatch bp, which requires one of the two cytosines in a base pair to be protonated at the N3 position (Figure 1E). Two C-rich duplexes intercalate between each other in an antiparallel manner such that consecutive C-C^+ motifs are placed almost at right angles to one another (Figure 1E). The i-motif quadruplex can be formed from four individual strands, although an intramolecular i-motif formed by folding of a single strand is biologically more relevant, owing to the common occurrence of ssDNA *in vivo* (during replication, in telomeres, etc). Lower pH, higher temperatures and high concentration of monovalent salts are known to favor the quadruplex form. Thus, at the near-neutral pH *in vivo* stability is probably attained in the presence of torsional stress or by inter-molecular interactions with proteins, RNA or drugs [81–83]. Stabilization of the i-motif has been associated with transcriptional regulation of gene activity [84]. Recently, the presence of i-motif *in vivo* has been demonstrated using antibodies that bind i-motifs with high selectivity and specificity [85]. Using immunostaining, the study showed that nuclei are especially rich in i-motif in the G1-phase of cell cycle, confirming the role of i-motif in transcriptional regulation and cell growth. As a result, a large fraction of i-motif research has focused on the folding pathways of C-rich telomeric single strand into the i-motif structure. Various experimental methods, including NMR spectroscopy, surface plasmon resonance (SPR), FRET, time-resolved (TR) infrared (IR) spectroscopy and stopped-flow circular dichroism (SFCD) and MD simulation techniques have been employed to unravel the complex dynamics of i-motif [86–89].

The folding pathway of intermolecular i-motif elucidated by measuring the association and dissociation rates of C-rich oligonucleotides using NMR spectroscopy revealed that under all conditions of temperature and pH, i-motif $[TC_n]_4$ formation follows a 3rd-order reaction kinetics [86]. The quadruplex formation occurs in a sequential manner by monomer strand association, with intermediate duplex and triplex states, with formation of either the triplex or quadruplex state being non-rate-limiting. SFCD analyses revealed that both folding and unfolding occur in the millisecond time scale (Figure 6A) [88]. By fitting the time evolution of the SFCD signal to a 1st order exponential decay function of the form $A + Be^{-k't}$, the folding pathway was deciphered to be a two-step process:

$$C m H^+ \underset{k_{-1}}{\overset{k_1}{\rightleftarrows}} C^* m H^+ \xrightarrow[n H^+]{k_2} I\,(m+n)\,H^+,$$

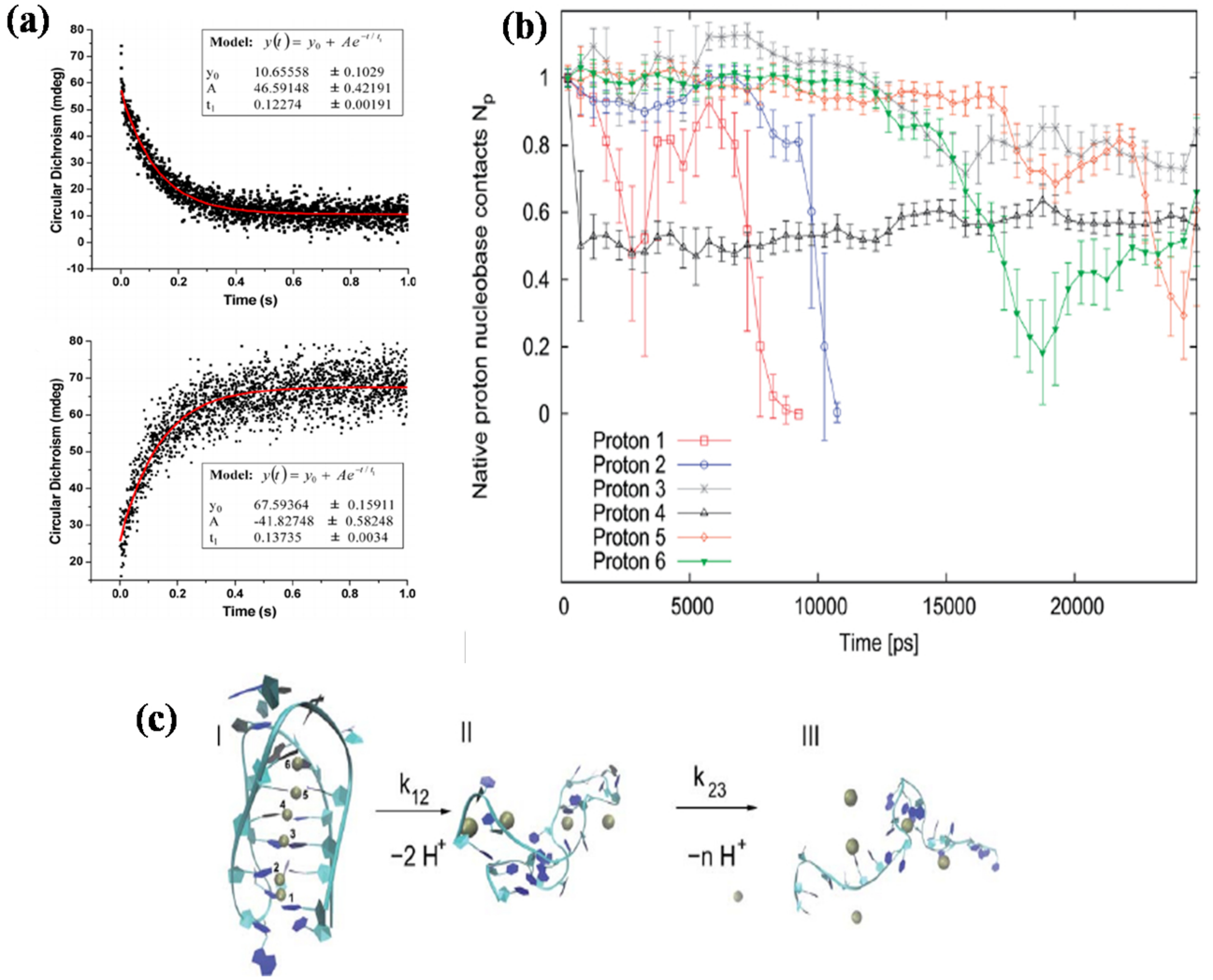

Figure 6. Folding pathway of the i-motif fold. (**A**) Intramolecular i-motif folding/unfolding kinetics was investigated using the SFCD technique by monitoring change in the CD band at 288 nm as a function of pH. Both unfolding (top panel) and folding (bottom panel) occur in the millisecond time scale. (Reprinted with permission from reference [88]. Copyright 2012 American Chemical Society). (**B**) Metadynamics simulations showed that these transient intermediated are formed by the loss of protons. (**C**) Structure I is the native i-motif, structure II illustrates hairpin configurations after proton loss whereas structure III represents an expanded and fully unfolded state (Reproduced from reference [89] with permission from The Royal Society of Chemistry).

where the ssDNA coil C (bound with *m* protons) first transits to an intermediate form C* in a pH-independent manner. Addition of at least three protons ($n \geq 3$) to C* occurs in a cooperative manner before C-C$^+$ bps can be formed. This step leads to an irreversible folding of the molecule to the i-motif (I). In agreement with these and other experimental studies, metadynamics analyses of an i-motif comprising of six C-C$^+$ bps showed that the 1st unfolding step occurs due to loss of two protons, leading to the formation of a hairpin (Figures 6B and C) [87, 89]. This is followed by the complete dissociation of the i-motif into a single-stranded random coil.

While the WC double helix remains the predominant structure *in vivo*, it can be induced to attain the quadruplex form. It has been shown that *in vitro* the duplex–quadruplex transition can be induced by changes in pH and temperature [2]. As shown in Figure 7, lower pH, higher temperatures and high concentration of monovalent salts favor the quadruplex form. In addition to the protective role of telomeric GQ described above, recent GQ imaging and mapping studies have also revealed the regulatory, and possibly therapeutic, role of GQ from non-telomeric regions in transcription and translation of certain proto-oncogenes. The role of GQ structures in replication and in modulating genome instability has also become very evident [30, 36].

Structural Polymorphism in Double-Stranded DNA

Evidently, the noncanonical forms of DNA are structurally versatile, physiologically significant and dynamically diverse. Nonetheless, the canonical B-form remains the most commonly occurring structure *in vivo* participating in all the major cellular processes that sustain life [90]. With the 1st crystal structure of a B-DNA dodecamer, it became evident that the B-DNA was not a rigid rod with fixed structural parameters [91, 92]. Conversely, it could more appropriately be defined as the 'B' family of DNA that encompasses diverse members, all complying with the overall geometries of the canonical WC model yet structurally and dynamically distinct from one another. Dynamic polymorphisms in the double helix encompass local

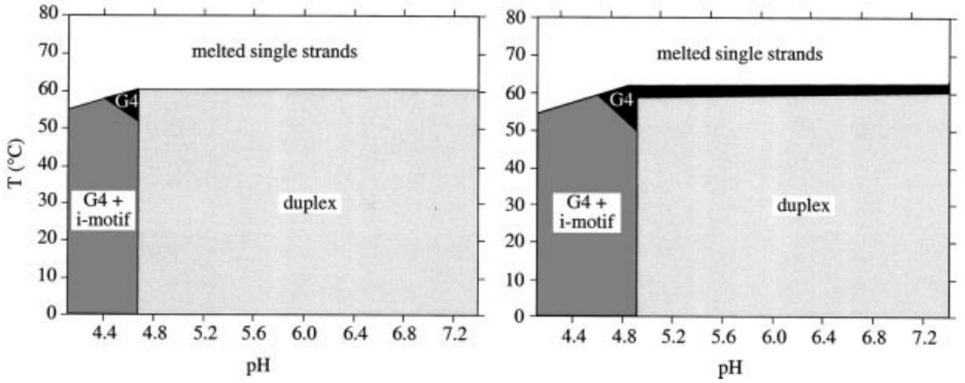

Figure 7. Duplex–quadruplex transitions under different environmental conditions are presented in the form of phase diagrams for the 1:1 mixture of oligonucleotides d[AGGG(TTAGGG)$_3$] and d[(CCCTAA)$_3$CCCT] in (**A**) NaCl and in (**B**) KCl. The diagram is divided into several areas, which correspond to melted strands (white), duplex (light gray) and quadruplex species (dark gray and black). The dark gray region corresponds to a domain where both quadruplexes (GQ and i-motif) coexist, whereas the black region corresponds to a domain where a folded GQ coexists with an unfolded C-rich strand [2].

fluctuations, like DNA breathing, bending, twisting, groove plasticity and base pair opening, as well as global deformations, like bending of DNA into supercoils and loops [93–96]. Structural and conformational dynamics exhibited by double-helical DNA is not only dependent on its sequence but also on the flanking nucleotides, the solvation environment and interacting partners [90, 97–101]. It is now known that such diversity in the shape and dynamics of similar DNA sequences largely guides specific recognition and binding of DNA motifs by proteins as well as small molecules [97–102].

Local dynamics of duplex DNA at femtosecond–nanosecond time scale has been probed by a wide range of spectroscopic techniques, while dynamics at the millisecond time scale has been obtained primarily from advanced NMR studies [53, 102, 103 and references therein]. Global deformations have been studied using single-molecule techniques, cyclization experiments [104], atomic force microscopy [105], molecular force sensors [106] and FRET [107]. With significant advances in atomistic force fields and computational power, microsecond-long state-of-the-art MD simulations are now possible and have contributed immensely to our understanding of motions in the B-DNA helix [103, 108–114]. Key findings obtained from these MD simulation studies will be reviewed here.

Local noncanonical dynamics of the canonical B-DNA

Considering the vast possibilities of different DNA sequences, several nucleic acid research groups across the globe constituted the Ascona B-DNA Consortium in 2002 to carry out large-scale MD studies on the double-helical B-DNA [115]. Microsecond-long MD simulations were carried out for 39 double-stranded 18-bp-long B-DNA oligomers, comprising of all 136 possible distinct tetranucleotide sequences [109]. In agreement with experimental observations, each oligomer presented multiple sequence-dependent substates within the average B-DNA structure. The study identified three conformational substates of B-DNA, defined in terms of three base pair step parameters—shift, slide and twist—that showed significant deviations from the average structure. The 1st substate is a characteristic of canonical B-DNA and was explored by all sequences. The 2nd is characterized by higher twist and negative shift and was largely restricted to RR steps (R = purine). The last region is characterized by very low twist and negative slide and was mainly populated by YR steps (Y = pyrimidine). Notably, RY steps appeared to be rigid and exhibited the sparsest dynamics. Structural and dynamic studies by our group and others have revealed that base pair step dynamics are correlated with conformational transitions in the sugar-phosphate backbone and variations in minor and major groove widths [110, 112–114, 116]. Spontaneous backbone transitions from BI/BI step to the mixed BI/BII or BII/BII state were commonly observed in all these studies, alongside a decrease in roll and increase in slide and twist (in BI, ε/ζ torsions are in t/g-; in BII, ε/ζ torsions are in g−/t conformation). Major and minor grooves were also found to exhibit significant plasticity with notable fluctuations in the groove width [103, 116].

Global deformation dynamics of the canonical B-DNA

Global deformations in DNA can be attributed to the intrinsic bendability of DNA. DNA can be predisposed to bending owing to inherent local properties of its base pair steps or induced to bend during supercoiling, packaging into nucleosomes or upon interaction with DNA binding proteins or ligands [108, 117]. Intrinsic DNA curvature was initially surmised to be dependent on the occurrence of periodic A-tract repeats due to their unusually large, negative propeller twist values. MD simulation studies have shown that the five-methyl group on thymine in A-tracts is also responsible for narrowing the minor groove and increasing curvature of A-tracts [118]. However, it is now known that several other sequences also exhibit significant curvature. A classic and widely investigated example of induced DNA bending is packaging of dsDNA over an octameric core of histone proteins to form nucleosome core particles (NCP). The dsDNA has to undergo a 7-fold compaction for packaging into chromatin, and yet maintain its overall B-DNA geometry, indicating the remarkable plasticity of DNA [119, 120]. Such strong bending induces sharp kinks in the DNA helix, which involves perturbation of base-stacking interactions and breakage of base pairs. MD simulation studies and single-molecule experiments have revealed that the nucleosomal DNA exhibits significant dynamics, which is key to the stability and functionality of the nucleosome. The three primary modes of DNA dynamics include two bending modes and a stretching mode. Once packed onto NCPs, the wrapped DNA fragment can bend further to maximize electrostatic interaction with the histone proteins [121] or unwrap transiently to expose the buried genetic information to gene regulatory proteins [122–124]. The wrapping/unwrapping dynamics of nucleosomal

DNA has been widely studied by single-molecule experiments. Single-molecule assays combining FRET and optical tweezers revealed that transient unwrapping of the core particle occurs at the stiffer end of the DNA, indicating that DNA flexibility is correlated with nucleosome stability [124]. The effect of modulations in DNA conformation on nucleosome disassembly was also shown by TR small-angle X-ray scattering and TR-FRET studies [125]. The time scale of such opening and closing of nucleosomal ends was found to occur at the rate of 0.1–1.0 ms^{-1} using single-molecule fluorescence correlation spectroscopic studies [126]. Further, large-scale analysis of nucleosomal DNA in crystal structures of NCPs in our laboratory revealed that structural versatility in DNA constituting similar nucleosome cores acts as a significant factor influencing the formation of nucleosomes in the vicinity of high-plasticity genes and in varying the probability of binding by regulatory proteins [120].

Summary

Structural and conformational polymorphs of canonical and noncanonical DNA are often located in key gene regulatory regions, like UTRs, promoters, telomeres, etc. and play a significant role in vital cellular processes. In these regions the physiologically predominant B-DNA attains the non-B-forms under certain environmental conditions, like presence of counter-ions or binding partners and superhelical stress. The dynamics of such transitions occur in the nanosecond–millisecond time scale and can now be investigated using advanced spectroscopic and computer simulation techniques. Advanced spectroscopic techniques employ probes to detect dynamical transitions induced by an external pulse e.g. by changing the temperature in laser temperature jump experiments or applying a radiofrequency field in carbon dispersion NMR techniques. Structural details of these dynamical states can be obtained using molecular simulation techniques. Single-molecule techniques have gone beyond ensemble-averaged properties to probe the heterogeneity of one molecule at a time. The current review is an attempt to emphasize how multiple physicochemical techniques used in conjunction can resolve previously intractable questions about the complex dynamics of noncanonical DNA structures.

Key Points

- The canonical double-helical DNA can attain noncanonical forms, which encompass lower- or higher-order structures like single-stranded hairpins, triplexes or quadruplex, like GQ and i-motif.
- The noncanonical DNA forms are often stabilized by non-WC base pairing, commonly HG or reverse HG hydrogen bonds.
- Within its conformational limits, the canonical B-DNA also undergoes extensive breathing dynamics at local base pair level as well as global deformations like bending and supercoiling.
- The dynamics of these transitions include folding/unfolding of B-DNA <—> noncanonical DNA forms, WC <—> HG transitions and few others, which have been studied using advanced spectroscopic and computer simulation techniques.
- Understanding the dynamics of noncanonical DNA and factors regulating their dynamics can further our understanding of their physiological roles.

Supplementary Data

Supplementary data are available online at http://bib.oxfordjournals.org/.

Acknowledgements

M.B. gratefully acknowledges Indian National Science Academy and Department of Science and Technology (Govt. of India) for fellowship. D.G. acknowledges Department of Science and Technology-Science and Education Research Board (Govt. of India) for fellowship.

References

1. Watson JD, Crick FHC. Molecular structure of nucleic acids. *Nature* 1953;**171**(5451):737–8.
2. Phan AT, Mergny J-L. Human telomeric DNA: G-quadruplex, i-motif and Watson-Crick double helix. *Nucleic Acids Res* 2002;**30**(21):4618–25.
3. Ghosh A, Bansal M. A glossary of DNA structures from A to Z. *Acta Crystallogr Sect D Biol Crystallogr* 2003;**59**(4):620–6.
4. Phan A-T, Kuryavyi V, Patel D. DNA architecture: from G to Z. *Curr Opin Struct Biol* 2015;**16**(3):288–98.
5. Jungkweon C, Tetsuro M. Conformational changes of non-B DNA. *Chem Soc Rev* 2011;**40**(12):5893–909.
6. Neidle S . *Principles of Nucleic Acid Structure*. Academic Press, 2008.
7. Olson WK, Colasanti AV, Lu X-J, *et al*. Physicochemical properties of nucleic acids: character and recognition of Watson-Crick base pairs. In: *Wiley Encyclopedia of Chemical Biology*. NY: John Wiley & Sons, 2007.
8. Palchaudhuri R, Hergenrother PJ. DNA as a target for anticancer compounds: methods to determine the mode of binding and the mechanism of action. *Curr Opin Biotechnol* 2007;**18**(6):497–503.
9. Seeman NC. Nanomaterials based on DNA. *Annu Rev Biochem* 2010;**79**:65–87.
10. Ren J, Wang J, Han L, *et al*. Kinetically grafting G-quadruplexes onto DNA nanostructures for structure and function encoding via a DNA machine. *Chem Commun* 2011;**47**(38):10563–5.
11. Umemura K. Hybrids of nucleic acids and carbon nanotubes for nanobiotechnology. *Nanomaterials* 2015;**5**(1):321–50.
12. Kawaguchi M, Yamazaki J, Ohno J, *et al*. Preparation and binding study of a complex made of DNA-treated single-walled carbon nanotubes and antibody for specific delivery of a 'molecular heater' platform. *Int J Nanomedicine* 2012;**7**:4363–72.
13. Goldman N, Bertone P, Chen S, *et al* . Toward practical high-capacity low-maintenance storage of digital information in synthesised DNA. *Nature* 2013;**494**(7435):77–80.
14. Saenger W. *Principles of Nucleic Acid Structure*. New York USA: Springer-Verlag, 1984;**49**:6221.
15. Leontis NB, Westhof E. Geometric nomenclature and classification of RNA base pairs. *RNA* 2001;**7**(4):499–512.
16. Hoogsteen K. The structure of crystals containing a hydrogen-bonded complex of 1-methylthymine and 9-methyladenine. *Acta Cryst* 1959;**12**:822–3.
17. Nikolova EN, Zhou H, Gottardo FL, *et al*. A historical account of Hoogsteen base-pairs in duplex DNA. *Biopolymers* 2013;**99**(12):955–68.

18. Abrescia NG, Thompson A, Huynh-Dinh T, *et al* . Crystal structure of an antiparallel DNA fragment with Hoogsteen base pairing. *Proc Natl Acad Sci* 2002;**99**(5):2806–11.
19. Ghosal G, Muniyappa K. Hoogsteen base-pairing revisited: resolving a role in normal biological processes and human diseases. *Biochem Biophys Res Commun* 2006;**343** (1):1–7.
20. Aishima J, Gitti RK, Noah JE, *et al*. A Hoogsteen base pair embedded in undistorted B-DNA. *Nucleic Acids Res* 2002;**30**(23):5244–52.
21. Zou L . Single- and double-stranded DNA: building a trigger of ATR-mediated DNA damage response. *GENES Dev* 2007;**21**(81):879–85.
22. Zou L, Elledge SJ. Sensing DNA damage through ATRIP recognition of RPA-ssDNA complexes. *Science* 2003;**300** (5625):1542–8.
23. Petruska J, Arnheim N, Goodman MF. Stability of intrastrand hairpin structures formed by the CAG/CTG class of DNA triplet repeats associated with neurological diseases. *Nucleic Acids Res* 1996;**24**(11):1992–8.
24. Chen X, Mariappan SV, Catasti P, *et al*. Hairpins are formed by the single DNA strands of the fragile X triplet repeats: structure and biological implications. *Proc Natl Acad Sci* 1995;**92** (11):5199–203.
25. Brázda V, Laister RC, Jagelská EB, *et al*. Cruciform structures are a common DNA feature important for regulating biological processes. *BMC Mol Biol* 2011;**12**:33–48.
26. Schleifman EB, Chin JY, Glazer PM. Triplex-mediated gene modification. *Methods Mol Biol* 2008;435:175–90.
27. Strobel SA, Dervan PD. Site-specific cleavage of a yeast chromosome by oligonucleotide-directed triple-helix formation. *Science* 1990;**249**(4964):73–5.
28. Strobel SA, Doucette-Stamm LA, Riba L, *et al*. Site-specific cleavage of human chromosome 4 mediated by triple-helix formation. *Science* 1991;**254**(5038):1639–42.
29. Buske FA, Bauer DC, Mattick JS, *et al* . Triplexator: detecting nucleic acid triple helices in genomic and transcriptomic data. *Genome Res* 2012;**22**(7):1372–81.
30. Rhodes D, Lipps HJ. Survey and summary of G-quadruplexes and their regulatory roles in biology. *Nucleic Acids Res* 2015;**43**(18):8627–37.
31. Lipps HJ, Rhodes D. G-quadruplex structures: *in vivo* evidence and function. *Trends Cell Biol* 2009;**19**(18):414–22.
32. Wright EP, Huppert JL, Waller ZAE. Identification of multiple genomic DNA sequences which form i-motif structures at neutral pH. *Nucleic Acids Res* **2016**, **45**:2951–9.
33. Mir B, Serrano I, Buitrago D, *et al* . Prevalent sequences in the human genome can form mini i-motif structures at physiological pH. *J Am Chem Soc* 2017;**139**(40):13985–8.
34. Gallego Â, Golden EB, Stanley DE, *et al* . The folding of centromeric DNA strands into intercalated structures: a physicochemical and computational study. *J Mol Biol* 1999;**285**(3):1039–52.
35. Bochman ML, Paeschke K, Zakian VA. DNA secondary structures: stability and function of G- quadruplex structures. *Nat Rev Genet* 2011;**21**:131–4.
36. Hänsel-Hertsch R, Di Antonio M, Balasubramanian S. DNA G-quadruplexes in the human genome: detection, functions and therapeutic potential. *Nat Rev Mol Cell Biol* 2017;**18**(5):279–84.
37. Capra JA, Paeschke K, Singh M, *et al* . G-quadruplex DNA sequences are evolutionarily conserved and associated with distinct genomic features in *Saccharomyces cerevisiae*. *PLoS Comput Biol* 2010;**6**(7):e1000861.
38. Huppert JL. Four-stranded nucleic acids: structure, function and targeting of G-quadruplexes. *Chem Soc Rev* 2008;**37**(7):1375–84.
39. Neidle S . Quadruplex nucleic acids as targets for anticancer therapeutics. *Nat Rev Chem* 2017;**1**:0041.
40. Martínez JM, Elmroth SK, Kloo L. Influence of sodium ions on the dynamics and structure of single-stranded DNA oligomers: a molecular dynamics study. *J Am Chem Soc* 2001;**123**(49):12279–89.
41. Bloomfield VA, Crothers DM, Tinoco I Jr. *Nucleic Acids: Structures, Properties, and Functions*. CA: University Science Books, 2000.
42. Smith SB, Cui Y, Bustamante C. Overstretching B-DNA: the elastic response of individual double-stranded and single-stranded DNA molecules. *Science* 1996;**271**(5250):795–9.
43. Chakraborty K, Mantha S, Bandyopadhyay S. Molecular dynamics simulation of a single-stranded DNA with heterogeneous distribution of nucleobases in aqueous medium. *J Chem Phys* 2013;**139**(7):075103.
44. Chakraborty K, Khatua P, Bandyopadhyay S. Exploring ion induced folding of a single-stranded DNA oligomer from molecular simulation studies. *Phys Chem Chem Phys* 2016;**18**(23):15899–910.
45. Ansari A, Kuznetsov SV, Shen Y. Configurational diffusion down a folding funnel describes the dynamics of DNA hairpins. *Proc Natl Acad Sci* 2001;**98**(14):7771–6.
46. Nayak RK, Peersen OB, Hall KB, *et al* . Millisecond time-scale folding and unfolding of DNA hairpins using rapid-mixing stopped-flow kinetics. *J Am Chem Soc* 2012;**134**(5):2453–6.
47. Portella G, Orozco M. Multiple routes to characterize the folding of a small DNA hairpin. *Angew Chem Int Ed* 2010;**49**(42):7673–6.
48. Nishigaki K, Kaneko Y, Wakuda H, *et al* . Type II restriction endonucleases cleave ssDNAs in general. *Nucleic Acids Res* 1985;**13**(16):5747–60.
49. Nishigaki K, Yuzuru Husumi A, Tsubota M. Detection of differences in higher order structure between highly homologous single-stranded DNAs by low-temperature denaturant gradient gel electrophoresis. *J Biochem* 1986;**99**(33): 663–71.
50. Orita M, Iwahana H, Kanazawa H, *et al* . Detection of polymorphisms of human DNA by gel electrophoresis as single-strand conformation polymorphisms. *Proc Natl Acad Sci* 1989;**86**(8):2766–70.
51. Nakabayashi Y, Nishigaki K. Single-strand conformation polymorphism (SSCP) can be explained by semistable conformation dynamics of single-stranded DNA. *J Biochem* 1996;**120**(2):320–5.
52. Larsen LA, Jespersgaard C, Andersen PS. Single-strand conformation polymorphism analysis using capillary array electrophoresis for large-scale mutation detection. *Nat Protoc* 2007;**2**(6):1458–66.
53. Sekhar A, Kay LE. NMR paves the way for atomic level descriptions of sparsely populated, transiently formed biomolecular conformers. *Proc Natl Acad Sci* 2013;**110**(32):12867–74.
54. Hansen AL, Nikolova EN, Casiano-Negroni A, *et al*. Extending the range of microsecond-to-millisecond chemical exchange detected in labeled and unlabeled nucleic acids by selective carbon R-NMR spectroscopy. *J Am Chem Soc* 2009;**131**(11):3818–9.
55. Alvey HS, Gottardo FL, Nikolova EN, *et al*. Widespread transient Hoogsteen base pairs in canonical duplex DNA with variable energetics. *Nat Commun* 2014;**5**:1–8.

56. Cubero E, Abrescia NGA, Subirana JA, *et al* . Theoretical study of a new DNA structure: the antiparallel Hoogsteen duplex. *J Am Chem Soc* 2003;**125**(47):14603–12.
57. Cubero E, Luque FJ, Orozco M. Theoretical study of the Hoogsteen-Watson-Crick junctions in DNA. *Biophys J* 2006;**90**(3):1000–8.
58. Sathyamoorthy B, Shi H, Zhou H., *et al* . Insights into Watson-Crick/Hoogsteen breathing dynamics and damage repair from the solution structure and dynamic ensemble of DNA duplexes containing m1A. *Nucleic Acids Res* 2017;**45**(9):5586–601.
59. Nikolova EN, Kim E, Wise AA, *et al* . Transient Hoogsteen base pairs in canonical duplex DNA. *Nature* 2011;**470**(7335):498–504.
60. Yang C, Kim E, Pak Y. Free energy landscape and transition pathways from Watson-Crick to Hoogsteen base pairing in free duplex DNA. *Nucleic Acids Res* 2015;**43**(16):7769–78.
61. Zhou H, Hintze BJ, Kimsey IJ, *et al* . New insights into Hoogsteen base pairs in DNA duplexes from a structure-based survey. *Nucleic Acids Res* 2015;**43**(7):3420–33.
62. Nair DT, Johnson RE, Prakash S, *et al* . Replication by human DNA polymerase-i occurs by Hoogsteen base-pairing. *Nature* 2004;**430**(6997):377–80.
63. Murat P, Balasubramanian S. Existence and consequences of G-quadruplex structures in DNA. *Curr Opin Genet Dev* 2014;**25**:22–9.
64. Ray A, Panigrahi S, Bhattacharyya D. A comparison of four different conformations adopted by human telomeric G-quadruplex using computer simulations. *Biopolymers* 2016;**105**(2):83–99.
65. Šponer J, Bussi G, Stadlbauer P, *et al* . Folding of guanine quadruplex molecules–funnel-like mechanism or kinetic partitioning? An overview from MD simulation studies. *Biochim Biophys Acta Gen Subj* 2017;**1861**(5 Pt B): 1246–63.
66. Šket P, Plavec J Tetramolecular DNA quadruplexes in solution: insights into structural diversity and cation movement. *J Am Chem Soc* 2010;**132**(36):12724–32.
67. Campbell NH, Parkinson GN Crystallographic studies of quadruplex nucleic acids. *Methods* 2007;**43**(4):252–63.
68. Joly L, Rosu F, Gabelica V. d(TGnT) DNA sequences do not necessarily form tetramolecular G-quadruplexes. *Chem Commun* 2012;**48**(67):8386–8.
69. Petraccone L, Spink C, Trent JO, *et al*. Structure and stability of higher-order human telomeric quadruplexes. *J Am Chem Soc* 2012;**133**(51):20951–61.
70. Bessi I, Jonker HRA, Richter C, *et al* . Involvement of long-lived intermediate states in the complex folding pathway of the human telomeric G-quadruplex. *Angew Chem Int Ed* 2015;**54**(29):8444–8.
71. Marchand A, Gabelica V. Folding and misfolding pathways of G-quadruplex DNA. *Nucleic Acids Res* 2016;**44**(22): 10999–1012.
72. Wang ZF, Li MH, Chen WW, *et al* . A novel transition pathway of ligand-induced topological conversion from hybrid forms to parallel forms of human telomeric G-quadruplexes. *Nucleic Acids Res* 2016;**44**(88):3958–68.
73. Hou X-M, Fu Y-B, Wu W-Q, *et al* . Involvement of G-triplex and G-hairpin in the multi-pathway folding of human telomeric G-quadruplex. *Nucleic Acids Res* 2017;**45**(19):11401–12.
74. Gray RD, Trent JO, Chaires JB Folding and unfolding pathways of the human telomeric G-quadruplex. *J Mol Biol* 2014;**426**(8):1629–50.
75. Stadlbauer P, Kührová P, Banáš P, *et al* . Hairpins participating in folding of human telomeric sequence quadruplexes studied by standard and T-REMD simulations. *Nucleic Acids Res* 2015;**43**(20):9626–44.
76. Štefl R, Cheatham TE, Špačková N, *et al* . Formation pathways of a guanine-quadruplex DNA revealed by molecular dynamics and thermodynamic analysis of the substates. *Biophys J* 2003;**85**(3):1787–804.
77. Bian Y, Tan C, Wang J, *et al* . Atomistic picture for the folding pathway of a Hybrid-1 type human telomeric DNA G-quadruplex. *PLoS Comput Biol* 2014;**10**(4):e1003562.
78. Mashimo T, Yagi H, Sannohe Y, *et al* . Folding pathways of human telomeric type-1 and type-2 G-Quadruplex structures. *J Am Chem Soc* 2010;**132**(42):14910–8.
79. Stadlbauer P, Krepl M, Cheatham TE, *et al* . Structural dynamics of possible late-stage intermediates in folding of quadruplex DNA studied by molecular simulations. *Nucleic Acids Res* 2013;**41**(14):7128–43.
80. Guéron M, Leroy J. The i-motif in nucleic acids. *Curr Opin Struct Biol* 2000; **10**(3):326–31.
81. Shafer RH, Smirnov I. Biological aspects of DNA/RNA quadruplexes. *Biopolymers* 2000;**56**(3):209–27.
82. Guo K, Pourpak A, Beetz-rogers K, *et al* . Formation of pseudo-symmetrical G-quadruplex and i-motif structures in the proximal promoter region of the RET oncogene. *J Am Chem Soc* 2008;**129**(33):10220–8.
83. Sun D, Hurley LH The importance of negative superhelicity in inducing the formation of G-quadruplex and i-motif structures in the c-Myc promoter: implications for drug targeting and control of gene expression. *J Med Chem* 2009;**52**(9):2863–74.
84. Kendrick S, Kang HJ, Alam MP, *et al* . The dynamic character of the BCL2 promoter i-motif provides a mechanism for modulation of gene expression by compounds that bind selectively to the alternative DNA hairpin structure. *J Am Chem Soc* 2014;**136**(11):4161–71.
85. Zeraati M, Langley DB, Schofield P, *et al* . I-motif DNA structures are formed in the nuclei of human cells. *Nat Chem* 2018;**10**(6):631–637.
86. Leroy JL. The formation pathway of i-motif tetramers. *Nucleic Acids Res* 2009;**37**(12):4127–34.
87. Keane PM, Wojdyla M, Doorley GW, *et al* . Long-lived excited states in i-motif DNA studied by picosecond time-resolved IR spectroscopy. *Chem Commun* 2014;**50**(23):2990–2
88. Chen C, Li M, Xing Y, *et al* . Study of pH-induced folding and unfolding kinetics of the DNA i-motif by stopped-flow circular dichroism. *Langmuir* 2012;**28**(51):17743–8.
89. Smiatek J, Heuer A. Deprotonation mechanism of a single-stranded DNA i-motif. *RSC Adv* 2014;**4**(33):17110–3.
90. Bansal M, Kumar A, Yella VR. Role of DNA sequence based structural features of promoters in transcription initiation and gene expression. *Curr Opin Struct Biol* 2014;**25**: 77–85.
91. Wing R, Drew H, Takano T, *et al* . Crystal structure analysis of a complete turn of B-DNA. *Nature* 1980;**287**(5784): 755–8.
92. Dickerson RE. The DNA helix and how it is read. *Sci. Am.* 1983;**249**(6):94–111.
93. Corless S, Gilbert N. Investigating DNA supercoiling in eukaryotic genomes. *Brief Funct Genomics* 2017;**16**(6): 379–89.
94. Cloutier TE, Widom J. DNA twisting flexibility and the formation of sharply looped protein-DNA complexes. *Proc Natl Acad. Sci* 2005;**102**(10):3645–50.

95. Travers AA . Why bend DNA? *Cell* 1990;**60**(22):177–80.
96. Matsumoto A, Olson WK. Sequence-dependent motions of DNA: a normal mode analysis at the base-pair level. *Biophys J* 2002;**83**(1):22–41.
97. Marathe A, Karandur D, Bansal M. Small local variations in B-form DNA lead to a large variety of global geometries which can accommodate most DNA-binding protein motifs. *BMC Struct Biol* 2009;**9**:24.
98. Yella VR, Bansal M. DNA structural features and architecture of promoter regions play a role in gene responsiveness of *S. cerevisiae*. *J Bioinform Comput Biol* 2013;**11**(6):1343001.
99. Rohs R, West SM, Liu P, *et al* . Nuance in the double-helix and its role in protein-DNA recognition. *Cur Opin Struct Biol* 2009;**19**(22):171–7.
100. Dror I, Golan T, Levy C, *et al* . A widespread role of the motif environment in transcription factor binding across diverse protein families. *Genome Res* 2015;**25**(9):1268–80.
101. Dror I, Rohs R, Mandel-Gutfreund Y. How motif environment influences transcription factor search dynamics: finding a needle in a haystack. *BioEssays* 2016;**38**(7): 605–12.
102. von Hippel PH, Johnson NP, Marcus AH. Fifty years of DNA 'breathing': reflections on old and new approaches. *Biopolymers* 2013;**99**(12):923–54.
103. Galindo-Murillo R, Roe DR, Cheatham TE. On the absence of intrahelical DNA dynamics on the μs to ms timescale. *Nat Commun* 2014;**5**:1–8.
104. Vafabakhsh R, Ha T. Extreme bendability of sub-100 bp long DNA revealed by single molecule cyclization. *Science* 2012;**337**(6098):1097–101.
105. Wiggins PA, van der Heijden T, Moreno-Herrero F, *et al* . High flexibility of DNA on short length scales probed by atomic force microscopy. *Nat Nanotechnol* 2006;**1**:137–41.
106. Shroff H, Reinhard BM, Siu M, *et al* . Biocompatible force sensor with optical readout and dimensions of 6 nm^3. *Nano Lett* 2005;**5**(7):1509–14.
107. Yuan C, Chen H, Lou XW, *et al* . DNA bending stiffness on small length scales. *Phys Rev Lett* 2008;**100**(1): 018102–018105.
108. Curuksu J, Zacharias M, Lavery R, *et al* . Local and global effects of strong DNA bending induced during molecular dynamics simulations. *Nucleic Acids Res* 2009;**37**(11): 3766–73.
109. Pasi M, Maddocks JH, Beveridge D, *et al* . μABC: a systematic microsecond molecular dynamics study of tetranucleotide sequence effects in B-DNA. *Nucleic Acids Res* 2014;**42**(19):12272–83.
110. Ben Imeddourene A, Elbahnsi A, Guéroult M, *et al* . Simulations meet experiment to reveal new insights into DNA intrinsic mechanics. *PLoS Comput Biol* 2015;**11**(12): 1–27.
111. Dans PD, Danilane L, Ivani I, *et al* . Long-timescale dynamics of the Drew-Dickerson dodecamer. *Nucleic Acids Res* 2016;**44**(9):4052–66.
112. Dršata T, Pérez A, Orozco M, *et al* . Structure, stiffness and substates of the Dickerson-Drew dodecamer. *J Chem Theory Comput* 2013;**9**(1):707–21.
113. Madhumalar A, Bansal M. Sequence preference for BI/BII conformations in DNA: MD and crystal structure data analysis. *J Biomol Struct Dyn* 2005;**23**(1):13–27.
114. Bhattacharyya D, Bansal M. Groove width and depth of B-DNA structures depend on local variation in slide. *J Biomol Struct Dyn* 1992;**10**(1):213–6.
115. Beveridge DL, Cheatham TE, Mezei M. The ABCs of molecular dynamics simulations on B-DNA, circa 2012. *J Biosci* 2012;**37**(3):379–97.
116. Pérez A, Luque FJ, Orozco M. Dynamics of B-DNA on the microsecond time scale. *J Am Chem Soc* 2007;**129**(47): 14739–45.
117. Kanhere A, Bansal M. An assessment of three dinucleotide parameters to predict DNA curvature by quantitative comparison with experimental data. *Nucleic Acids Res* 2003;**31**(10):2647–58.
118. Marathe A, Bansal M. The 5-methyl group in thymine dynamically influences the structure of A-tracts in DNA at the local and global level. *J Phys Chem B* 2010;**114**(16): 5534–46.
119. Richmond TJ, Davey CA. The structure of DNA in the nucleosome core. *Nature* 2003;**423**(6936):145–50.
120. Marathe A, Bansal M. An ensemble of B-DNA dinucleotide geometries lead to characteristic nucleosomal DNA structure and provide plasticity required for gene expression. *BMC Struct Biol* 2011;**11**:1–21.
121. Pasi M, Lavery R. Structure and dynamics of DNA loops on nucleosomes studied with atomistic, microsecond-scale molecular dynamics. *Nucleic Acids Res* 2016;**44**(11):5450–6.
122. Li G, Levitus M, Bustamante C, *et al* . Rapid spontaneous accessibility of nucleosomal DNA. *Nat Struct Mol Biol* 2005;**12**(1):46–53.
123. Böhm V, Hieb AR, Andrews AJ, *et al* . Nucleosome accessibility governed by the dimer/tetramer interface. *Nucleic Acids Res* 2011;**39**(8):3093–102.
124. Yodh JG, Ha T, Ngo TTM, *et al*. Asymmetric unwrapping of nucleosomes under tension directed by DNA local flexibility. *Cell* 2015;**160**(6):1135–44.
125. Chen Y, Tokuda JM, Topping T, *et al* . Asymmetric unwrapping of nucleosomal DNA propagates asymmetric opening and dissociation of the histone core. *Proc Natl Acad. Sci* 2017;**114**(2):334–9.
126. Wei S, Falk SJ, Black BE. A novel hybrid single molecule approach reveals spontaneous DNA motion in the nucleosome. *Nucleic Acids Res* 2015;**43**(17):e111.
127. Case DA, Cerutti DS, Cheatham TE, III, *et al*. *AMBER 2017*. *San Francisco*: University of California, 2017.