

*Review*

## **Making the Bend: DNA Tertiary Structure and Protein-DNA Interactions**

**Sabrina Harteis and Sabine Schneider \***

Department of Chemistry, Technische Universität München, Lichtenbergstrasse 4,  
85747 Garching, Germany; E-Mail: sabrina.harteis@gmail.com

\* Author to whom correspondence should be addressed; E-Mail: sabine.schneider@mytum.de;  
Tel./Fax: +49-89-289-13336.

*Received: 21 May 2014 / in revised form: 1 July 2014 / Accepted: 1 July 2014 /*

*Published: 14 July 2014*

---

**Abstract:** DNA structure functions as an overlapping code to the DNA sequence. Rapid progress in understanding the role of DNA structure in gene regulation, DNA damage recognition and genome stability has been made. The three dimensional structure of both proteins and DNA plays a crucial role for their specific interaction, and proteins can recognise the chemical signature of DNA sequence (“base readout”) as well as the intrinsic DNA structure (“shape recognition”). These recognition mechanisms do not exist in isolation but, depending on the individual interaction partners, are combined to various extents. Driving force for the interaction between protein and DNA remain the unique thermodynamics of each individual DNA-protein pair. In this review we focus on the structures and conformations adopted by DNA, both influenced by and influencing the specific interaction with the corresponding protein binding partner, as well as their underlying thermodynamics.

**Keywords:** DNA structure; DNA structure recognition; protein-DNA interactions

---

### **1. The Overall Topology of DNA**

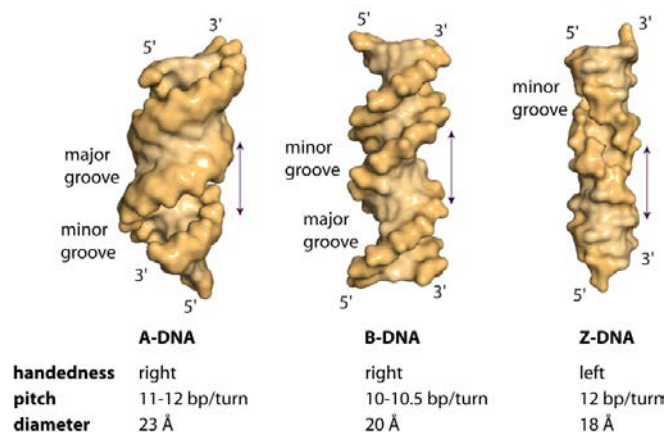
The three different general topologies taken up by the DNA double helix are termed A, B and Z. These conformations are distinguished by the handedness of the helix, their pitch (the distance between base and the base obtained after a full 360° turn), the number of nucleotides within one pitch and the distance between consecutive bases (the rise) [1]. Figure 1 schematically depicts the physical as well

as the geometric parameters defining the location and orientation of the base pair steps relative to each other in the DNA helix. While RNA double strands all exist in the A-conformation, DNA adopts mainly the B-conformation (Figure 2a,b), with both forming right-handed helices. However, DNA double strands are able to take up the A-conformation in some protein-DNA complexes and under dehydrated conditions [2,3]. DNA-RNA hybrid duplexes adopt conformations between A- and B-type duplex geometry, with the RNA strand having an overall A-type structure and the DNA strand a structure intermediate of A- and B-type [4]. Depending on the nucleic acid sequence and binding partner, conformational changes to either A- or B-type duplexes are possible [5,6]. In the A-form helix, the ribose is puckered in the C3' endo, which in dsRNA is caused by the steric hindrance of the C2' endo puckering by the sugar 2'hydroxyl group. In contrast, in the B-form, C2' endo ribose is found. This difference in sugar puckering results in a reduced P-P distance of 5.9 Å in the A-conformation compared to 7.0 Å in the B-conformation, as well as a shortening of the distance between the stacked bases (B-conformation: 3.30–3.37 Å; A-conformation: 2.59–3.29 Å). In addition, the A-form helix is slightly unwound with 11–12 nucleotides for every 360° turn, while in B-DNA there are 10–10.5 nucleotides. The helix axis runs almost straight through the centre of the base pair in the B-form, while in the A-form the centre of the base pairs are shifted about 4.5 Å from the axis. Thus the B-helix features two grooves, the major and the minor grooves, which differ in their width but are equally deep. (Figure 2) In contrast, the A-form helix possesses a small but deep major groove, only accessible to water and metal ions, and a shallow, but wide minor groove. (Figure 2) The structural properties of the DNA grooves strongly influence the recognition and interaction with the protein partners that will be discussed below.

The term Z-DNA stems from the observed zig-zag conformation of the phosphate backbone of a left-handed helix taken up by alternating purine-pyrimidine DNA sequences (GC repeats) under high salt conditions [7]. Here, due to the displacement of the base pairs away from the axis, only one groove can be observed that is analogous to the minor groove of B-DNA. The bases forming the major groove in B-DNA are reorganized in Z-DNA as such that they build a convex outer surface. (Figure 2c) Each guanine base is rotated around the glycosidic bond into *syn*-conformation, with the sugar puckered in C3' endo-conformation, while the cytosine in the adjacent base step is in the C2' endo, *anti*-conformation. Compared to A- and B-DNA, Z-DNA possesses a helical diameter of 18 Å, with 12 bp per helical turn, a rise of 3.7 Å and rotation by 30° per bp. In addition, the phosphate groups are closer together. Due to the electrostatic repulsion of the phosphate groups and the energy penalty associated with rotation of the Gs into *syn*-conformation, under physiological conditions the Z-conformation is the less favoured higher energy state and the DNA is pushed into the B-form. This also explains why Z-DNA becomes the stable conformation under high-salt concentrations, since the salt decreases the electrostatic repulsion of the phosphate [1,8].



**Figure 2.** Surface representation of A-, B- and Z-DNA. Hexadecameric A- and B-DNA helices with alternating (ATGC)<sub>4</sub> sequence were generated in COOT [13]. Z-DNA with the sequence (GC)<sub>8</sub> was constructed with 3DNA [14]. A-DNA is slightly unwound, possessing a shallow, wide minor groove and a deep, narrow major groove. B-DNA features a narrow minor groove and a broad major groove. In the left-handed Z-DNA, named by the zig-zag pattern in the phosphodiester backbone, a deep minor groove is observed, while the major groove is reorganised and features a convex surface.

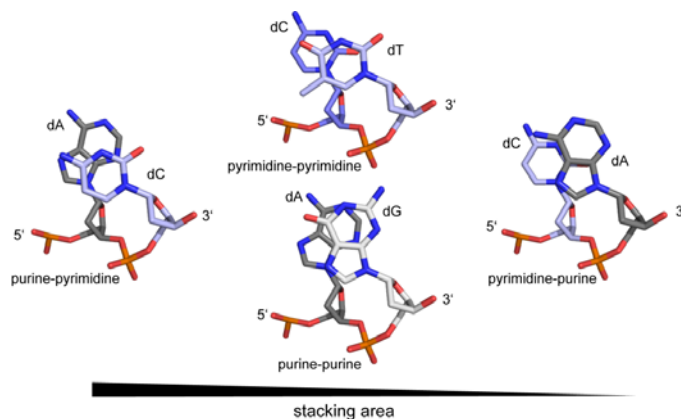


## 2. DNA Sequence and DNA Structure

In recent years, it became apparent that the genome contains additional information to the triplet genetic code. Within and outside protein coding regions, overlapping to the genetic code, DNA shape plays a crucial role in gene regulation, genome organisation and integrity [15–17]. It was postulated that proteins recognise and use the intrinsic flexibility of the DNA [18,19]. Nevertheless, the sequence of the DNA is an essential prerequisite that a certain structure can be adopted. Here, the electronic configurations of the base pairs, the number of H-bonds and the presence of exocyclic groups in the major and minor grooves not only determine the deformability of the DNA, but also the deformation energy necessary to adopt a particular conformation [20–22]. For example, the melting temperature of any given DNA duplex is dependent on the number of H-bonds in the base pairs (GC *versus* AT). Moreover the sequence also influences the persistence length (the quantitative term for the stiffness of a polymer) of the DNA, which is directly linked to the base pair rigidity. A/T rich sequences have a lower persistence length and are therefore more bendable than G/C rich sequences [23]. Another factor impacting the DNA rigidity is the stacking interactions of the base pairs themselves. A purine-pyrimidine base step is thermodynamically more stable due to the larger stacking area than purine-purine or pyrimidine-pyrimidine base steps, with the pyrimidine-purine being the least stable. (Figure 3) Thus the persistence length of the DNA is determined by the sequence as well as its composition and context, with the stacking energy being the major factor for the base step stability. The deformability of a given DNA strand, which is defined as the range of conformational space that can be adopted, is inversely related to its persistence length. Thus sequences with high melting and stacking energy (AC/GT, GC and GA/TC) are on average less deformable compared to sequences with low melting and stacking energy (*i.e.*, AA/TT, AT and TA) [24–27]. Investigations on the flexibility of DNA strands with biological relevant length (5–100 nm~15–300 bp) by atomic-force microscopy have shown that the

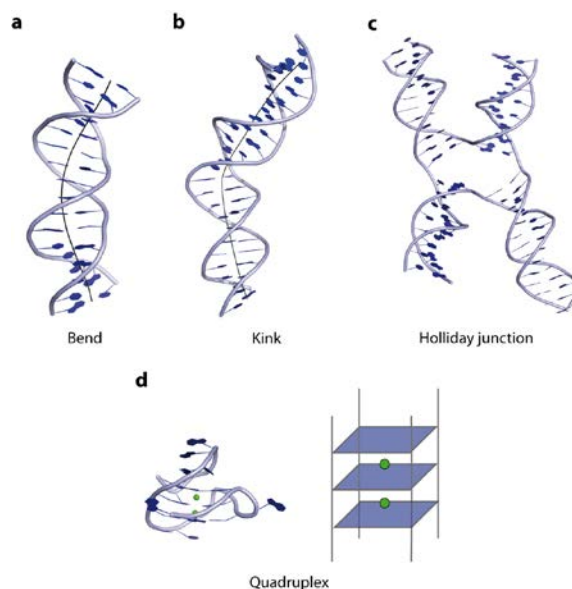
probability of spontaneous sharp bending is higher than predicted. Thus the bendability cannot be extrapolated from long-length-scale measurements and local effects like the DNA sequence must be taken into account [28]. Nevertheless electrostatic and nonelectrostatic effects play a role in determining the stiffness of DNA [29]. In addition, geometric analysis of available DNA structures showed that on a base-step level many conformers exist that cannot be classified as either A- or B-type, and/or represent an A-to-B transitional state [30].

**Figure 3.** Direct correlation of the stacking area of base steps in DNA and their stiffness.



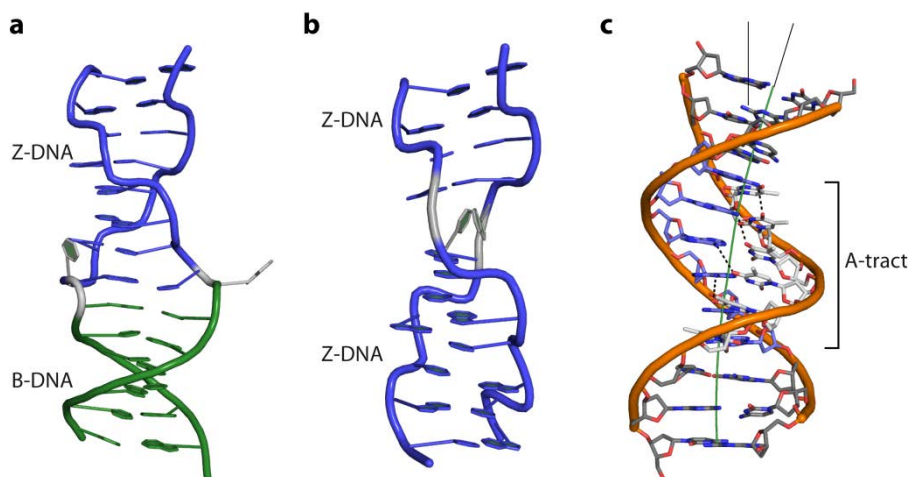
In the following paragraphs, we illustrate the structural variations taken up by the DNA, like Z-DNA conformation, bends, kinks and intriguing tertiary higher order structures, such as quadruplexes and cruciforms, and their underlying sequence dependency (Figure 4).

**Figure 4.** Illustration of DNA tertiary structures. (a) Curvature of the double helix over several bases results in a DNA bend (PDB code 1JJ4); (b) A DNA kink causes a change of strand orientation in an otherwise linear double strand (PDB code 2KEI); (c) Holliday junctions are formed by strand exchange between two DNA double helices (PDB code 2QNC); (d) Quadruplex DNA, a four-stranded structure, consists of guanine rich sequences, harbouring metal ions (highlighted in green) in the centre (PDB code 3QXR).



As already mentioned above, GC repeats undergo a B- to Z-conformation transition in the presence of high salt concentrations. Next to the structural B-to-Z transition point, the so-called B-Z-junction, the helical parameter corresponds to standard B- and Z-DNA (Figure 5a). Interruption of the dinucleotide repeats by single base insertions or deletions brings neighbouring helices out of phase, and a Z-to-Z junction is formed, requiring less energy than a B-Z-junction [31] (Figure 5b). In genomes long stretches of GC repeats are rare since they represent hot spots of instability [32]. *In vivo* Z-conformations of DNA are found in regions of supercoiled B-DNA near the promoter regions where they stimulate transcription [1,33,34]. It was shown that DNA is negatively supercoiled in Z-conformation when it becomes unwrapped from histones. This prevents reformation of the nucleosome, since Z-DNA cannot form nucleosomes and the site remains free for transcription factors to bind and initiation of transcription [8,35]. Moreover, during transcription the DNA behind the moving RNA polymerase is unwound and subjected to negative torsional strain [36] further stabilising Z-DNA formation near the transcription start site. A number of proteins specifically interacting with Z-DNA are known to date: the fish kinase PKZ, the innate immune system receptor ZBP1 (also known as DLM-1 and DA), the pox-virus inhibitor of interferon (IFN) response E3L and the IFN-induced form of the RNA editing enzyme ADAR1. They are all involved in processes related to the IFN system, important for defending the organism against viruses, microbes and tumour cells [37–41].

**Figure 5.** (a) Cartoon representation of a BZ-junction (PDB code 2ACJ) and (b) a Z-Z junction (PDB code 3IRR); (c) Intrinsic curvature of an A-tract containing DNA (PDB code 1D98) due to the propeller twist of its base pairs. Non-Watson-Crick H-bonds, between the base steps in the major groove, which rigidify the DNA strand, are shown as dashed lines.



Bent DNA is curved over a stretch of several bases resulting in different orientation of the regions on both sides of the curvature (Figure 4a). It was shown that short repeated stretches of poly-TA sequences or repetitive runs of 4–6 adenine base pairs (“A-tracts”) introduce an intrinsic curvature of the helix [27]. At the same time they increase the rigidity of the DNA strand, which contradicts the observation that AT-rich sequences are generally more flexible. The crystal structure of a poly-TA DNA explains these observed physical properties: the bases within a base pair are not coplanar but possess a propeller twist and are additionally buckled (Figure 5c). This results in a reduced helical

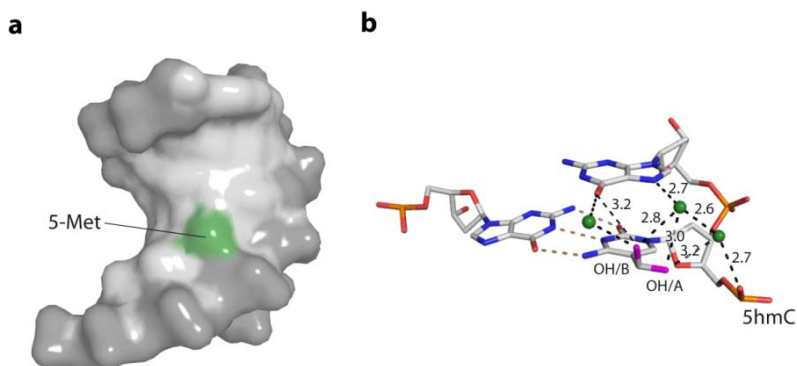
repeat (10 base pairs per turn) and narrowing of the minor groove [42–44]. The structure is further stabilized and hence rigidified by non-Watson-Crick H-bonds formed down the major groove, enabled by the propeller twist of the base pairs. Bending of DNA containing such poly-TA stretches occurs due to their zero net roll. In standard B-DNA the net roll angle of each half turn are not zero and cancel each other out, resulting in a straight strand. Thus if an AT-stretch is present, the roll angle is not cancelled out causing a change in direction of the helix at the junction to the AT-stretch. (Figure 5c) [23] The increased stiffness prevents wrapping of such sequences around histones. In yeast, runs of about 20 bp of such poly-TA sequences are found upstream from promoter elements of constitutively active transcribed genes [45,46]. Moreover, the presence of guanine bases in the minor groove prevents bending of DNA due to the strong steric constraint imposed by the exocyclic 2-amino group [22]. In contrast, a kink in the DNA is defined as local unstacking of a single base pair step in an otherwise linear DNA strand, causing a change of orientation of the helix (Figure 4b). In this case the pyrimidine-purine (TA; CA) steps exhibiting the lowest stacking energy appear to act as a flexible “hinge” in DNA-protein interactions (Figure 3) [47–49].

Beyond the canonical bases, epigenetic DNA base modifications, such as the methylation of cytosines (5mC) alter the anisotropic DNA bendability or flexibility at the particular site and can be linked to their regulatory effects. It was argued that the presence of 5mC rigidifies the backbone of the DNA [50] increases base pair stacking [51,52] and alters the solvation dynamics in the major groove [53]. It was also shown that 5mC changes the net curvature of A-tract containing DNA [54]. Ultimately, the presence of 5mC in CpG-islands of promoter regions affects not only the protein-binding partners [55], but also the wrapping around histones results in changes of histone positioning [21,56,57]. Recently, the influence of 5mC and hydroxymethyl C (5hmC) in DNA duplexes on the structure of the DNA and the interaction with the basic helix-loop-helix (bHLH) transcription factors Max and USF was investigated. This revealed that while no direct structural impact of the modification on the B-DNA was observed, the symmetrical presence on both DNA strands of either modified base completely abolished protein interaction, while hemi-modification was partly tolerated by the transcription factors [58]. Figure 6 shows the structural basis for the altered bendability of 5mC- and 5hmC-containing DNA, by reaching of the 5-methylgroup in the major groove of the DNA, altering the charge distribution and steric properties. Interestingly, 5mC in context of a GC repeat region facilitates the B- to Z-DNA transition [59].

The formation of higher order tertiary structures such as junctions, cruciforms and quadruplexes strongly depend on the DNA sequence. Cruciform structures, also known as Holliday junctions (Hj), occur when the four helices of two DNA duplexes are interconnected by strand exchange at a branch point, forming a 4-way, stacked X-shaped structure [60–63] (Figure 4c). All Hjs consist of a stem, a branch point and a loop, with the loop size depending on the length of the gap between the repeats. For their formation either perfect or imperfect inverted repeats of 6 bases or more are required. The gap sequence was shown to have a direct impact on Hj formation, with AT-rich sequences possessing the highest probability. In the absence of  $Mg^{2+}$  ions, Hjs have a square-planar conformation, while in their presence conformational isomer forms of stacked-X junctions are observed, all of which are in rapid exchange with each other. Nevertheless the equilibrium for a distinct conformer is biased not only by metal ions but also by the DNA sequence [64,65].



**Figure 6.** Structural bases for the increased stiffness of DNA containing the epigenetic bases 5mC and 5hmC. **(a)** The 5-methylgroup of 5mC (green) reaches in the major groove of the DNA, altering the charge distribution and steric properties in the major groove; (PDB code 4C63) **(b)** Direct and water-mediated hydrogen bonding of the two alternative hydroxyl conformations (pink) of 5hmC to the 3' base and phosphate backbone. Water molecules are depicted as green spheres. (PDB code 4C5X)



Quadruplexes, also termed tetraplexes or G4 structures, are four-stranded DNA structures, formed by guanine-rich sequences. Here, Hoogsteen-hydrogen-bonded guanines build up tetrads in a cation dependent manner [66] (Figure 4d). *In vivo*, sequences giving rise to quadruplexes have been identified in G-rich eukaryotic telomeres and promoter regions [67–70]. It was recently shown that the formation of quadruplexes in eukaryotes is modulated in the course of DNA replication [71]. In addition, sites containing G-quadruplex structures are particularly prone to DNA strand breaks and chromosomal rearrangements [72,73]. A number of proteins interacting with G-quadruplexes have been identified, all of which are involved in genome maintenance such as helicases (FANCI, BLM), nucleosome remodelling (ATXR) or DNA damage tolerance (Rev1) [74–77]. Their proposed functions are the prevention of genetic and epigenetic instability at G-quadruplex sites, control of telomere length, as well as resolving quadruplex structures formed during DNA replication [78]. Unfortunately, to date, no structural data on quadruplexes bound by their protein partner are available.

### 3. Thermodynamic Consideration of DNA-Protein Interactions

In order to discuss the influence of DNA structure and sequence on the specific binding with a protein, interaction thermodynamics need to be considered. A protein has to bind significantly more to the recognition site than to competing, non-specific DNA. The equilibrium constant for non-specific binding ranges from  $10^3$ – $10^6$   $M^{-1}$  per site, with a free energy change of about  $-4$  to  $-7$  kcal/mole per site. Specific binding is 1000 to  $>10^7$  times tighter, but the affinity must not be excessive ( $<10^{12}$  per mole per site) since the binding needs to be reversible [79–81]. The gain in free energy in specific protein-DNA interactions is about  $-11$  kcal/mole and per site. Chemical reactions and interactions only take place if the free energy  $\Delta G_0$  is negative. Simplified, the free energy is the difference between the changes in enthalpy and entropy. Factors contributing positively to the change in enthalpy upon interaction are the formation of salt-bridges, non-polar contacts and hydrogen bonds. Nevertheless, the interaction has an entropic cost due to the loss of translational, rotational and configurational freedom



and vibrations of entrapped waters. Enthalpically disadvantageous is the desolvatisation of the interface [82]. The net gain in enthalpy is further reduced by bond bending and unfavourable non-bonded interaction. Particularly, positioning of functional groups, residues and bases not at their lowest level of potential energy are disadvantageous. In addition, bending, distortion or destacking of the DNA increases the enthalpic cost. For instance, the TATA binding protein (TBP) unstacks 6 bases when binding to its respective promoter region, which costs 50–60 kcal/mol [83]. Generally, favourable changes in enthalpy are accompanied by an entropic penalty, while on the contrary favourable changes in entropy carry an enthalpic cost. For instance little or no distortion of the DNA is good for the enthalpic cost, but is unfavourable for the entropy change. The release of water from polar surfaces increases the entropy at an enthalpic cost. The transfer of a water molecule from bulk solution to a specific site carries an entropic penalty, which needs to be balanced by favourable van der Waals and hydrogen bonding interactions. Thus for the free energy to be negative, favourable enthalpy changes need to drive unfavourable entropy changes and *vice versa* [83,84].

#### 4. Base Readout, Pre-Shaped DNA and Protein Recognition

DNA is a polyelectrolyte with high axial charge density and many counter ions/or water molecules at its surface. Therefore DNA binding proteins are often characterised by positively charged amino acids, such as arginine or lysine residues at the binding site for the interactions with the negative charged phosphate backbone. There are two mechanisms by which proteins specifically recognise a particular DNA: the base readout and the shape readout (reviewed in [85]). In the base readout mechanism, the specificity is achieved through direct or water mediated contacts with the DNA bases in the major and minor grooves. The major groove has the highest potential for base readout recognition since the functional groups of the four bases are displayed and are therefore accessible to specific interactions with the protein partner (A: 6-NH<sub>2</sub>, 7-N; C: 4-NH<sub>2</sub>; G: 6-O, 7-N; T: 4-O, 7-CH<sub>3</sub>). In contrast, only a two-letter code can be read in the minor groove (A: 3-N; C: 2-O; G: 3-N, 2-NH<sub>2</sub>; T: 2-O) (Figure 6). Therefore some sequences cannot be distinguished in the minor groove [86,87]. One also needs to take into account that bidentate H-bonds convey higher selectivity than single H-bonds [88]. In a number of protein-DNA complex structures, base specificity is mediated by conserved water molecules, which can be regarded as non-covalent extensions of the DNA bases [82]. An example for the use of waters in sequence-specific recognition is the Trp repressor-operator complex, which will be discussed below.

However, the sequence specificity for the majority of DNA binding proteins cannot be explained solely on the basis of specific protein-base contacts. For instance, transcription factors (TFs) achieve *in vivo* unique sequence specificity with seemingly identical DNA motifs but distinct targets. In a recent study Gordan *et al.* [89] demonstrated that the *Saccharomyces cerevisiae* bHLH TFs, Cbf1 and Tye1, bind their recognition sequence (E-box motif) depending on their genomic context. Further computational analyses suggest that nucleotides outside the E-box motifs contribute to specificity by influencing the three-dimensional structure of the DNA binding sites. Thus proteins evidently utilise the additional information from DNA structure and DNA deformability. In the example mentioned above, the bHLH TFs recognise the specific bases of the conserved E-box in the major groove,

whereas local DNA shape recognition in the flanking regions appears to enable distinct DNA binding preferences among paralogous TFs [16,90].

The role played by the DNA sequence in this so-called shape readout mechanism can be explained through its influence on the conformational space predominantly occupied by a particular DNA as already described above. In this case a certain structure/conformation can be recognised and its stabilisation in the complex with the protein is thermodynamically favourable [91]. The shape readout mechanism can be further differentiated as local and global shape readout [85], depending whether the DNA deviation from B-DNA is local (kink) or more general (bend, A- or Z-DNA). However, one has to take into account that variation of DNA shape always influences the base readout: conformational changes of the DNA alter the geometry and hence accessibility of the bases in the major and minor grooves. Figure 7 highlights differences in the major and minor groove and therefore base readout options of B-DNA, bent DNA as well as kinked DNA.

In the following section, the recognition mechanisms will be illustrated in more detail, using structural examples of DNA-protein complexes. Therefore, the examples have been divided into: base readout (restriction endonuclease *HindIII*); base readout combined with intrinsic shape readout (restriction endonuclease *EcoRV*, Trp-repressor protein); sequence-context dependent base readout (LexA repressor); intrinsic shape readout (TATA binding protein); and overall shape readout (chromosomal protein HU, T7 endonuclease, ADAR1).

**Figure 7.** Base readout in the major and minor groove. (a) Functional groups of the DNA base pairs in the major and minor DNA groove; (b) Accessibility of functional groups in the major groove in a 16-mer B-DNA; (c) Locally increased readout accessibility in minor groove of bent B-DNA (PDB code 1JJ4); (d) Enlarged binding capability due to kink-induced shape alteration in B-DNA (PDB code 2KEI, [92]). Hydrogen bond donors in blue, acceptors in red and thymine methyl group in green.

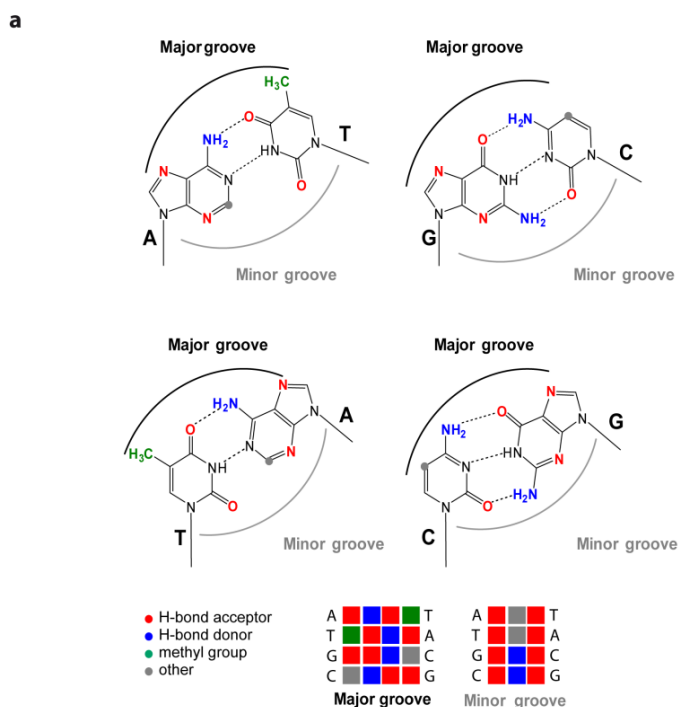
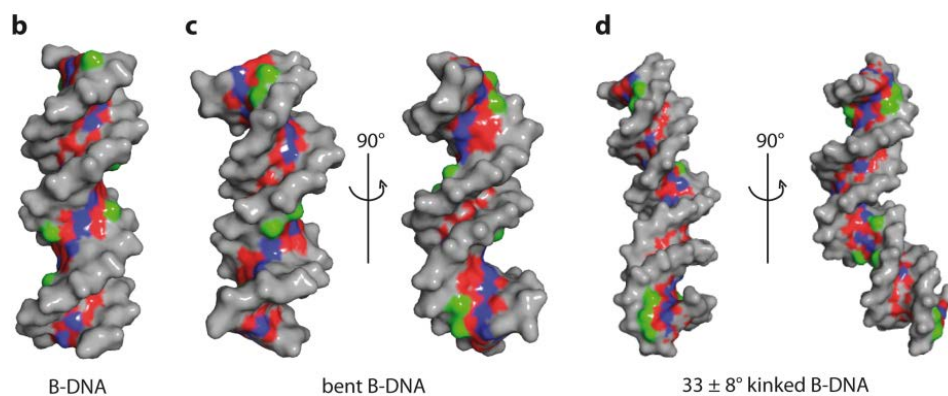


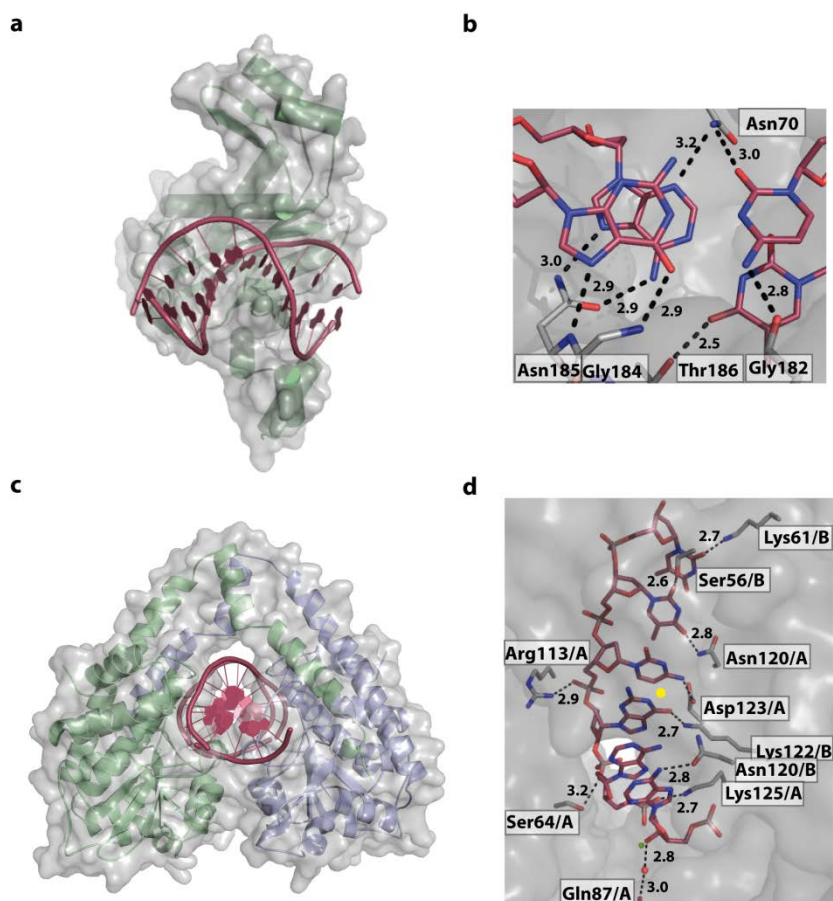
Figure 7. Cont.



#### 4.1. Base Readout and Recognition Sequence Intrinsic Shape Readout—The Restriction Endonucleases *HindIII* and *EcoRV*

In the last century the discovery of bacterial restriction endonucleases has revolutionized molecular biology and numerous studies on their structures, mechanisms of action and DNA recognition are available to date. Bacteria have evolved restriction endonucleases as weapons against foreign DNA, such as viruses like bacteriophages [63,93,94]. These enzymes must be highly specific to cleave only DNA containing the respective recognition site, without degrading the host genomic DNA. Thus restriction endonucleases are an excellent, well investigated example for specific DNA sequence recognition in combination with enzymatic activity. They are grouped into four classes, according to: the nature of their target sequence; the cleavage position relative to the recognition sequence; co-factor; and structure. Type II restriction endonucleases cut DNA within or close to their palindromic recognition site [95], which they need to bind independent of the sequence context. Examples for classical type II endonucleases are *HindIII* from *Haemophilus influenzae* and *EcoRV* from *E. coli*, which digest DNA at 5' A/AGCTT 3' and 5' GAT/ATC 3' sites, respectively. The structures of these enzymes in complex with their corresponding target DNA reveal that while *HindIII* relies predominately on direct base readout, *EcoRV* combines base- and local shape readout. Typical for type II endonucleases, two proteins matching the two fold rotational symmetry at the centre of the palindromic target sequence are bound to the DNA [96,97] (Figure 8). In the DNA-*HindIII* structure, both DNA strands and all six target bases located in the major groove are read out by the enzyme through 10 mono- and bidentate contacts, either directly or mediated by water molecules (Figure 8).

**Figure 8.** The restriction endonucleases *EcoRV* and *HindIII*. (a) Overall structure of the *EcoRV* DNA-complex (PDB code 1RVB). For clarity only one of two symmetrically binding protein molecules is shown. The DNA is kinked by 50°; (b) *EcoRV* interacts intensively with the outer GC and AT bases forming 8 direct contacts with almost all functional groups; (c) Structure of *HindIII*/DNA (PDB code 3A4K) complex displaying the symmetrical bound dimer embedding the DNA; (d) Active site of *HindIII* with, for clarity, one of the palindromic target sequence AAGCTT. Direct and water mediated contacts are formed. Water molecules, and for catalysis, required  $Mg^{2+}$  ions are shown as red and green spheres, respectively. The symmetry axis of the complex is located in middle of target sequence, between bases GC/CG and indicated as a yellow dot.



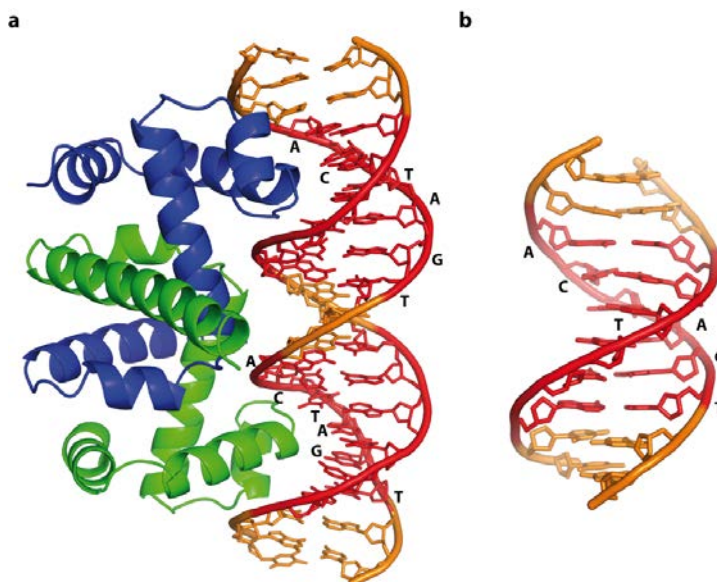
Two  $Mg^{2+}$ -ions are located in the active site for catalysis. In addition, the structure clearly shows why methylation by *H. influenza* adenine methylases [98,99] protects against cleavage by *HindIII*: conversion of the first A of the target sequence to 6-met-A inhibits the interaction of Asn 120 with the 6-N amino group. In comparison *EcoRV* interacts intensively with the outer GC and AT bases forming 8 direct contacts with almost all functional groups [97] (Figure 8a,b). Outside the recognition motifs, both *HindIII* and *EcoRV* interact with the sugar-phosphate backbone. Common to all restriction endonucleases, they initially bind to DNA non-specifically and when their recognition site is encountered, enzyme and DNA undergo an induced-fit conformational change, resulting in DNA cleavage and release. The first weak non-specific DNA binding releases water and counterions from the DNA-protein interface. This balances the thermodynamically unfavourable loss of translational and

rotational entropies upon complex formation that is further aided through the contacts made with the phosphate backbone, adding favourable enthalpy changes. Encounter with the cognate sequence triggers a cooperative conversion of the non-specific to the specific complex and couples sequence recognition to catalytic cleavage [95]. The DNA bound by *Hind*III is slightly bent by about 26° in comparison to ideal B-DNA [96] (calculated with the program CURVES [44]). The target motif with the GC base step centrally located in the major groove reduces the local flexibility. This is due to the stiffness of the stacking purine-pyrimidine moieties as well as the steric hindrance by the 6-O of the G, reaching in the major groove. In contrast, the central, flexible AT base steps of the *Eco*RV target sequence (5' GAT/ATC 3') allows kinking by 50°. In summary, *Hind*III uses base readout on its whole target sequence, accompanied with a slight bend of the DNA. In comparison, *Eco*RV combines base readout by directly contacting a part of the target sequence with readout of the local shape provided by the flexible central AT sequence [100]. The induced-fit conformational change with the distortion of the DNA was shown to be essential for catalytic cleavage by type II restriction endonucleases. In structures of non-specific complexes neither the phosphate, or the catalytic residues nor the Mg<sup>2+</sup> ions are positioned for cleavage [101]. Only upon DNA bending and structural changes of the enzymes the catalytic apparatus is assembled.

#### 4.2. Combining Base and Shape Readout—The *Escherichia coli* Trp Operator

One of the first examples where structural data on a DNA sequence alone and in complex with a protein were available is the *E. coli* trp operator DNA and the Trp repressor protein [91,102]. This allowed dissecting the impact from the DNA sequence and its resulting structure on the specific recognition by its protein-binding partner. Comparing the structure of the free DNA and the DNA-protein complex revealed that specificity is conferred cooperatively by direct base readout, structural effects imposed by the DNA sequence and a large contribution by water-mediated interactions. Ten water binding sites in the major groove are conserved in the free and bound DNA of which three mediate specific contacts between the protein and nitrogen atoms of the purine bases. Replacing the purine nitrogens with a carbon atom resulted in a free energy difference of 1 kcal/mol on average, clearly showing that these water-mediated interactions are critical for the formation of a high-affinity complex [103]. While the free DNA is straight, the operator DNA in the protein complex is bent by 15°, resulting in compression of the major groove and widening of the minor groove. However both free and bound DNA are slightly unwound with 10.6 bases per turn and possess a deeper major groove, due to displacement of the base pairs from the helix axis. The observed displacement values of 0.9–1.9 Å are between A- and B-form DNA, albeit the backbone parameters are typical for B-DNA (Figure 9a). The structure adopted by the free trp operator DNA largely resembles its bound form even though it might be influenced by crystal packing forces (Figure 9b). The deformation energy barrier is therefore likely to be low, further facilitated by the central, more flexible T-A step [91,102]. This example clearly shows that both DNA shape driven by its sequence as well as specific base—amino acid interactions determine the specificity.

**Figure 9.** Free trp operator DNA, and in a complex with the Trp repressor protein, illustrating direct base readout and its impact on DNA structure. Cognate recognition sequences are highlighted in red. (a) Schematic representation of the trp operator-Trp repressor complex (PDB code 1TRO); The 15° bend compresses the major groove in the complex, whereas free DNA (b) (NDB code BDJ061) is straight, highlighting the subtle differences between bound and free DNA.



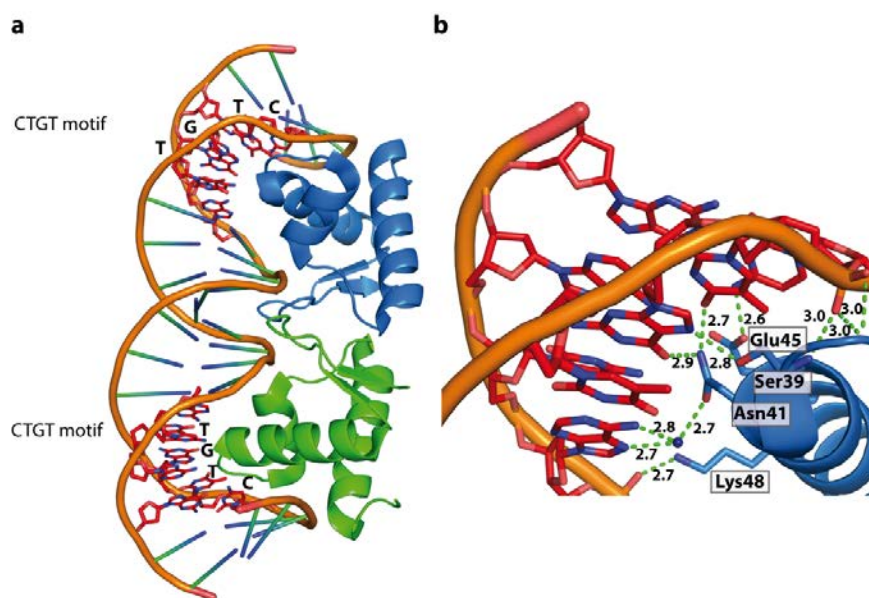
#### 4.3. Base Readout and DNA Shape Context—The *Escherichia coli* LexA Repressor

The LexA repressor recognises the conserved CTGT motif (SOS-box), heavily relying on base readout [104]. However, the overlaying code of sequence-dependent DNA shape and flexibility by the flanking and interspacing sequences modulate binding affinity or determine whether a DNA-LexA complex is formed. The LexA repressor can therefore serve as a general model of how proteins achieve unique sequence specificity with seemingly identical DNA motifs, but distinct *in vivo* targets. In the LexA-DNA complex extensive direct or water-mediated contacts between the protein and the CTGT motif in the major groove are observed (Figure 10). The major groove is bent towards the protein resulting in an overall 35° curvature of the bound DNA. The entropic cost associated with DNA bending is compensated by contacts between the protein and the phosphate backbone in the adjacent regions. The impact of the flanking regions flexibility on the binding affinity has been shown, and the highest affinities were reported when the consensus sequence is flanked by TA and AT base steps (5'-TACTGT(AT)<sub>4</sub>ACAGTA-3') [105,106]. Due to the high flexibility of the AT steps, they can be placed in the narrow minor groove in a bent complex structure, since they provide only a low entropic penalty. Thus the LexA repressor clearly binds to its recognition motif by a combination of base-readout, where the affinity is driven by shape-readout of the flanking regions. This is further demonstrated by binding studies where the AT-spacer was replaced by less flexible DNA sequences such as A-tracts or G-C sequences, resulting in marked reduction of the binding affinity [104]. In a cellular context, decreased binding affinity caused by properties of the flanking regions is sufficient to determine whether or not a complex is formed. At first glance, the LexA and the Trp repressor appear to operate with a comparable recognition mechanism. However, while the flexibility of the flanking



regions around the SOS-Box determines LexA binding, the Trp-operator sequence intrinsically encodes the DNA conformation recognised by the repressor protein.

**Figure 10.** LexA repressor dimer bound to its cognate SOS boxes (PDB code 3J5O). **(a)** Schematic representation of LexA/DNA complex; **(b)** Extensive direct or water mediated contacts between the winged DNA binding domain and CTGT motif result in a 35° DNA bend towards the major groove.

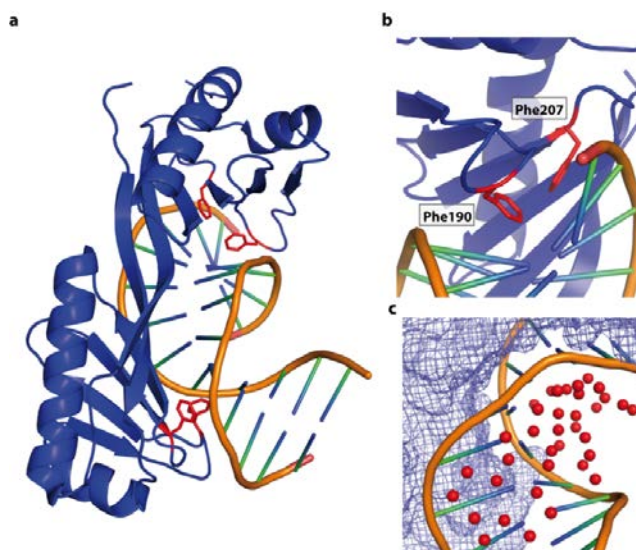


#### 4.4. Sequence Specific Shape Readout—The TATA Binding Protein and the TATA Box

The previous examples illustrate that specificity can be governed by base readout, while the flexibility intrinsic to the target DNA or the sequence context impacts on the affinity or enables binding. In contrast, recognition of the TATA-box by the TATA binding protein (TBP) is predominately driven by the adopted shape of the DNA. The TBP plays a crucial role in transcription initiation by recognising the TATA-Box found in the  $-10$  promoter region of many prokaryotic and eukaryotic promoters [107–110]. The structure of the heptameric TATA box recognition sequence bound to TBP (PDB code 1YTB) shows that the protein only contacts the minor DNA groove and no base specific contacts occur. The DNA in the complex is heavily unwound ( $105^\circ$  over 7 bp), which is compensated by a superhelical turn of  $120^\circ$ . The minor groove is widened, a feature similar to A-DNA. The smooth overall  $80^\circ$  bend ( $90^\circ$  in the central 6 bases) of the DNA is due to the additive effect of the large positive role angle of about  $26^\circ$  and the reduced twist of each base pair (Figure 11a). The compression of the major groove is stabilised by an extensive H-bond network and water molecules [111] (Figure 11c). In addition, pairs of phenylalanines penetrate the base pairs producing DNA kinks (Figure 11b). This indicates that the specificity is predominately driven by DNA shape [112]. The enthalpic cost for the bending of this DNA is lowered by the higher flexibility due to the reduced stacking of the T-A steps in the TATA-Box sequence, as observed for the trp operator. Further compensation is provided by the stabilisation through water-mediated and electrostatic interactions. Nevertheless the sequence flanking the TATA-box will have an impact on the overall stiffness and conformational space occupied by the DNA and hence on the binding affinity.



**Figure 11.** DNA shape recognition of the TATA box by the TBP binding (PDB code 1YTB). (a) Overall view of the complex. TBP interacts with the minor groove, resulting in a compressed major groove. The contacts are dominated by van der Waals' interactions occurring between nonpolar and polar atoms. The DNA is heavily unwound ( $105^\circ$ ) and displays two sharp kinks (highlighted in red); (b) Penetration of the double helix by Phe 190 and Phe 207 at the kink site; (c) The water molecules (red spheres) stabilize the major groove of the DNA. The protein surface is shown as blue mesh.



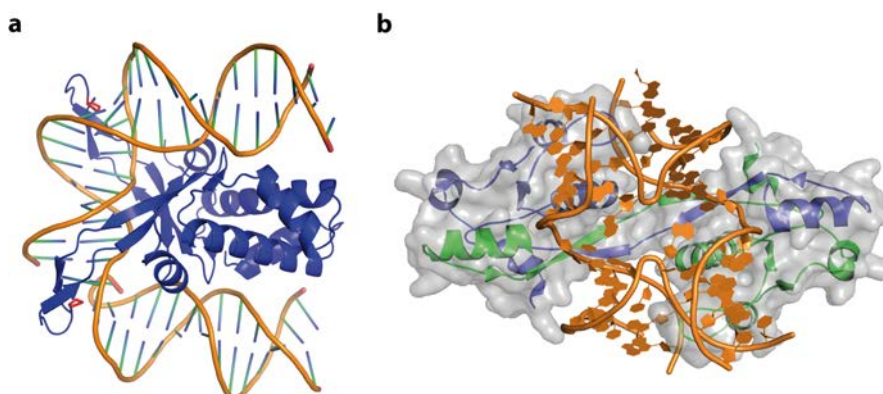
#### 4.5. DNA Shape Recognition

##### 4.5.1. The Bacterial Chromosomal Proteins HU and IHF

Further examples for DNA sequence driving DNA shape, with the resulting conformers being recognised rather than the bases by specific contacts, are the bacterial proteins HU (histone like protein from *E. coli* strain U93) and IHF (integration host factor). However the concept of sequence-dependent shape recognition in these examples is generalised, allowing the compensation for sequence variations. In the bacterial cell they are responsible for compacting the bacterial chromosome and maintenance of supercoiling and play an important role in DNA damage recognition, regulation of transcription and DNA replication [113–120]. Functionally, they are therefore homologous to the eukaryotic HMG box proteins [121,122]. HU and IHF have apparently no specific DNA target sequence, however it has been reported that HU, as well as IHF, bind tighter to A-tract [123], AT-rich sequences [124], pre-bent DNA, and DNA containing nicks or kinks [123,124]. This sequence preference was observed in a number of structural and biochemical studies on the protein-DNA complexes, which led to its thermodynamic explanation: both proteins bend the DNA about  $105^\circ$ – $180^\circ$ , base pairs within a 9 nucleotide distance are destacked, and the DNA is kinked twice. This base-destacking is accomplished by intercalating two conserved proline residues between the base pairs. No contacts with the major groove are observed and the DNA shape that is recognised has a large twist angle at the second dinucleotide step [125]. For IHF it could be shown that bending and tight specific binding is concerted [126], and it forces the two bends of the DNA to be almost coplanar (Figure 12a). The bend angles observed in HU-DNA complexes are less dramatic than in IHF-DNA complexes. However, the DNA has a negative

writhe and is undertwisted by 2 bp compared to IHF (3 bp compared to B-DNA) [124]. Thermodynamically, the observed bending, twisting and kinking generates a large, unfavourable strain. DNA kinking alone was estimated to cost 14.1 kcal/mol [127]. How do IHF and HU compensate for this energetic penalty? Isothermal titration calorimetry (ITC) showed that binding of IHF to DNA is enthalpy-driven through the reorganisation and formation of surface salt bridges between protein and DNA [128]. In addition, binding to pre-bent DNA, which can be caused by DNA nicks and lesions or AT-rich sequences, correlates with a narrow minor groove that fits better in the protein clamp, at a lower enthalpic cost. Nevertheless, both proteins compensate the cost for nonspecific DNA binding and kinking by increasing the length of the binding site and formation of more salt bridges [124,129], a clear demonstration of the entropy–enthalpy compensation. This principle also applies to wrapping of DNA around nucleosomes in eukaryotic genome organisation events. Eukaryotic nucleosomes bind predominantly to so-called “positioning sequences” and with particular high affinity to the artificial “601” sequence [130]. This naturally deformable 601 sequence is a unique nucleotide repeating pattern: five TA base-pair steps (TA-TA dimers) roughly recurring in phase with every double-helical turn (~10 bp), are alternated with A + T-rich and G + C-rich motifs at the half-helical turns [131].

**Figure 12.** DNA-binding by IHF and T7 endonuclease I. **(a)** In the IHF-DNA complex (PDB code 1IHF) the DNA is roughly bent, with proline residues (red) intercalating, destacking the base pairs and stabilising the kinked structure. No direct or indirect contacts of IHF with the DNA major groove can be observed; **(b)** Illustration of a Holliday junction bound by T7 Endonuclease I (PDB code 2PFJ). The protein-DNA interactions results in large conformational changes: the duplex arms of the HJ are kinked by 80° and the junction centre is opened.



#### 4.5.2. The Holliday Junction and T7 Endonuclease I

*In vivo* Hjs are key intermediate of homologous recombination and DNA double-strand break repair. Their presence in negative supercoiled DNA was demonstrated to be essential for efficient plasmid replication, as well as genome stability [73,132–134]. Moreover, strong evidence linking transcriptional regulation and HJ DNA in promoter and enhancer regions has been found [135,136]. Numerous proteins involved in replication (helicases [137]), transcription (14-3-3 proteins [138–140]), DNA repair (XPG and XPF [141–143]), and chromatin remodelling (HMG box protein family [144–146]), are known to interact with HJ DNA. The proteins binding Hjs are diverse in their mechanisms and

functions and have been reviewed extensively [147–152]. Structural data on H<sub>j</sub> alone and in complex with their protein-binding partner are available [153–155]. In the next paragraph we want to focus on the role played by the H<sub>j</sub> in the recognition event.

A well investigated example is the phage T7 endonuclease I (Endo I), which not only recognises and resolves H<sub>j</sub>s but also branched DNA structures and, less efficiently, single-base heteroduplexes [156]. The efficiency was demonstrated to depend on the conformer population distribution. Nevertheless, it was shown that T7 Endo I only binds tightly to branched DNA duplexes where the duplex arms can adopt an 80° angle. When the crystal structures of the junction and the enzyme alone are compared with the complex, large-scale structural changes on both macromolecules can be seen [157]. In the junction, the B-form arms are still aligned coaxial, however the handedness of the junction is changed by a rotation of 130°, resulting in the 80° angle and an opened junction centre (Figure 12b). Extensive interactions of the junction backbone, in particular the arms that are subject to cleavage, with the basic protein surface are observed [157]. The enzyme does not directly interact with the DNA bases, albeit DNA sequence preferences in the junction arms exist [158]. The flexible N-terminal region (residues 1–16), which is not resolved in the crystal structure, was shown to stabilise the transition state to the active complex by locally opening the junction structure at the strand exchange points, thus reducing the activation energy by 3.8 kcal/mol [159].

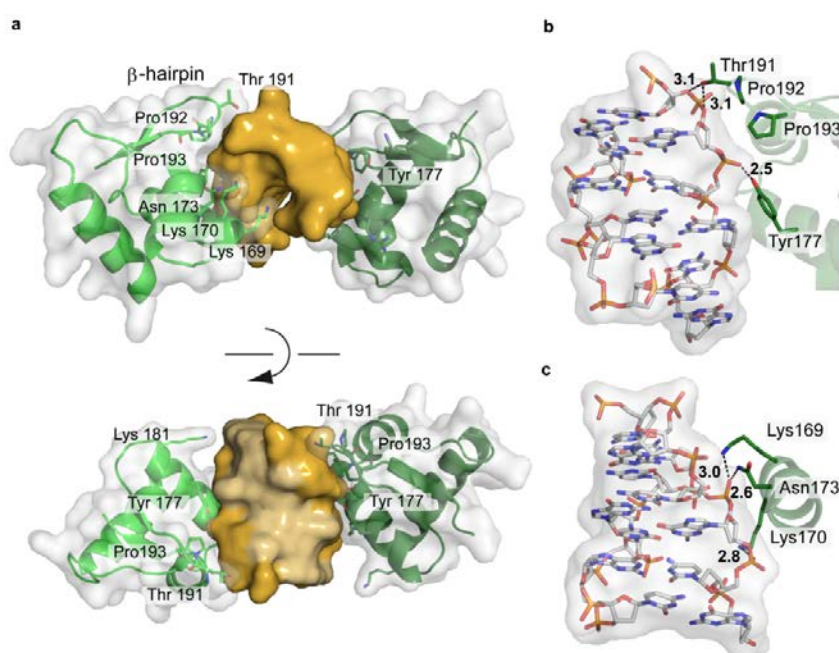
Are there common principles how proteins interact with H<sub>j</sub>? Generally they vary in their folding topology as well as the structure of the bound DNA. However, common to all H<sub>j</sub> resolving enzymes is that they themselves show no direct sequence readout, but form extensive interactions with the sugar-phosphate backbone and appear to recognise the overall shape and conformational space occupied by H<sub>j</sub>s in an induced-fit mechanism. Their dynamic character is crucial to allow moulding of the H<sub>j</sub>s onto the generally large binding surface of predominately dimeric protein binding partners, at little energetic cost, ultimately leading to the observed tight binding ( $K_D \sim 1$  nM) [158]. Since the same DNA junction can be bound by different enzymes, adopting distinct structures, this begs the question whether the protein binding partner imposes a shape upon the X-stacked structure, or traps a transient conformer of the free junction DNA.

#### 4.5.3. Z-DNA and ADAR1

The IFN-induced form of the RNA editing enzyme ADAR1 deaminates adenine in pre-mRNA to inosine, which codes as guanine. However, its N-terminal Z $\alpha$ -domain is responsible for high-affinity binding to Z-DNA [160,161]. A biological function of Z-DNA binding by Z $\alpha$  has not been clearly defined yet. However Z-DNA is stabilized by negative supercoiling, which is formed transiently upstream of an active RNA polymerase [36]. Thus ADAR1 might be recruited to actively transcribed genes to act upon the nascent RNA. In addition, it was shown that ADAR1 edits viral genomes during viral transcription and alters viral growth [162]. The interaction between ADAR1 and Z-DNA is driven solely by the shape of the left-handed DNA, with the Z $\alpha$ -domain's binding interface being complementary to the DNA in terms of conformation and electrostatic nature (Figure 13a). The Z $\alpha$ -domain has a HTH folding topology often found in proteins binding B-DNA. Nevertheless, the recognition modes are distinct, reflecting the different topologies of the DNA binding partner. While most HTH proteins use their helix  $\alpha_3$  to interact with the major groove of B-DNA, helix  $\alpha_3$  in ADAR1 only

contacts the outer surface of the Z-DNA. The C-terminal  $\beta$ -hairpin of ADAR1 largely contributes to the binding [163]. The protein does not contact the DNA bases themselves, but forms an extensive hydrogen bonding network, both direct and water-mediated, to five consecutive backbone phosphates [163,164] (Figure 13b,c). In addition, van der Waals interactions between the DNA and cis-proline residues, located at the  $\beta$ -hairpin tip, and aromatic tyrosine side chains can be observed. (Figure 13). Thus a general DNA-recognising protein folding topology is adapted to match the shape of its DNA binding partner.

**Figure 13.** Structure of the Z $\alpha$ -domain of ADAR1 bound to left-handed Z-DNA (PDB code 3IRQ). (a) Schematic representation of the protein-DNA complex, highlighting the surface fit of the two macromolecules. The DNA is shown in gold, with the two bound proteins in green, overlaid with their semi-transparent surface; (b,c) Electrostatic interactions between ADAR1 and the Z-DNA backbone.



## 5. Conclusions

Rapid progress has been made in understanding how DNA structure functions as an overlaying code to the DNA sequence and its role in gene regulation, DNA damage recognition and genome stability. Biochemical, biophysical and structural studies on DNA and DNA-protein complexes have provided penetrating insights into how DNA sequence impacts on the structural and physical properties of this macromolecule and hence enables or prevents protein recognition. On the molecular level, whether a DNA protein complex is formed is determined by its free energy and the enthalpic and entropic gain and cost associated with each particular interaction. In general the DNA sequence determines and enables not only distinct interactions, but also the overall conformational space occupied by the DNA and therefore its shape. In complexes predominately driven by base specific readout, the DNA deviates little from its free conformation. In contrast, when there are few or no base-specific interactions between protein and DNA, one can observe that proteins recognise and stabilise DNA shapes, strongly

diverging from the classical A- or B-DNA. Thus the finely balanced entropy—enthalpy compensation necessary for every interaction will be either rendered favourable and stable or unfavourable and instable.

### Acknowledgments

We would like to thank Markus Müller and Matthew Nodwell for critical reading of the manuscript. This work was supported by the German Research Foundation (DFG) and the Technische Universität München within the funding programme Open Access Publishing. We also thank the excellence cluster CIPSM and the Fond der chemischen Industrie for financial support.

### Author Contributions

SH and SS conducted the literature survey; SH prepared the figures; SS wrote the manuscript.

### Conflicts of Interest

The authors declare no conflict of interest.

### References

1. Dickerson, R.E. DNA structure from A to Z. *Methods Enzymol.* **1992**, *211*, 67–111.
2. Jacobo-Molina, A.; Ding, J.; Nanni, R.G.; Clark, A.D.; Lu, X.; Tantillo, C.; Williams, R.L.; Kamer, G.; Ferris, A.L.; Clark, P.; *et al.* Crystal structure of human immunodeficiency virus type 1 reverse transcriptase complexed with double-stranded DNA at 3.0 Å resolution shows bent DNA. *Proc. Natl. Acad. Sci. USA* **1993**, *90*, 6320–6324.
3. Lu, X.J.; Shakked, Z.; Olson, W.K. A-form conformational motifs in ligand-bound DNA structures. *J. Mol. Biol.* **2000**, *300*, 819–840.
4. Petersen, M.; Bondensgaard, K.; Wengel, J.; Jacobsen, J.P. Locked nucleic acid (LNA) recognition of RNA: NMR solution structures of LNA:RNA hybrids. *J. Am. Chem. Soc.* **2002**, *124*, 5974–5982.
5. Ban, C.; Ramakrishnan, B.; Sundaralingam, M. A single 2'-hydroxyl group converts B-DNA to A-DNA: Crystal structure of the DNA-RNA chimeric decamer duplex d(CCGGC)r(G)d(CCGG) with a novel intermolecular G C base-paired quadruplet. *J. Mol. Biol.* **1994**, *236*, 275–285.
6. Chen, X.; Ramakrishnan, B.; Sundaralingam, M. Crystal structures of B-form DNA-RNA chimeras complexed with distamycin. *Nat. Struct. Biol.* **1995**, *2*, 733–735.
7. Wang, A.H.; Quigley, G.J.; Kolpak, F.J.; Crawford, J.L.; van Boom, J.H., van der Marel, G.; Rich, A. Molecular structure of a left-handed double helical DNA fragment at atomic resolution. *Nature* **1979**, *282*, 680–686.
8. Rich, A., Zhang, S. Timeline: Z-DNA: The long road to biological function. *Nat. Rev. Genet.* **2003**, *4*, 566–572.
9. Dickerson, R.E. Definitions and nomenclature of nucleic acid structure components. *Nucleic Acids Res.* **1989**, *17*, 1797–1803.

10. Olson, W.K.; Bansal, M.; Burley, S.K.; Dickerson, R.E.; Gerstein, M.; Harvey, S.C.; Heinemann, U.; Lu, X.J.; Neidle, S.; Shakked, Z.; *et al.* A standard reference frame for the description of nucleic acid base-pair geometry. *J. Mol. Biol.* **2001**, *313*, 229–237.
11. Sinden, R.R. *DNA Structure and Function*; Academic Press: New York, NY, USA, 1994.
12. Rao, A.N.; Grainger, D.W. Biophysical properties of nucleic acids at surfaces relevant to microarray performance. *Biomater. Sci.* **2014**, *2*, 436–471.
13. Emsley, P.; Lohkamp, B.; Scott, W.G.; Cowtan, K. Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.* **2010**, *66*, 486–501.
14. Lu, X.J.; Olson, W.K. 3DNA: A versatile, integrated software system for the analysis, rebuilding and visualization of three-dimensional nucleic-acid structures. *Nat. Protoc.* **2008**, *3*, 1213–1227.
15. Kitayner, M.; Rozenberg, H.; Rohs, R.; Suad, O.; Rabinovich, D.; Honig, B.; Shakked, Z. Diversity in DNA recognition by p53 revealed by crystal structures with Hoogsteen base pairs. *Nat. Struct. Mol. Biol.* **2010**, *17*, 423–429.
16. Rohs, R.; West, S.M.; Sosinsky, A.; Liu, P.; Mann, R.S.; Honig, B. The role of DNA shape in protein-DNA recognition. *Nature* **2009**, *461*, 1248–1253.
17. Parker, S.C.; Hansen, L.; Abaan, H.O.; Tullius, T.D.; Margulies, E.H. Local DNA topography correlates with functional noncoding regions of the human genome. *Science* **2009**, *324*, 389–392.
18. Vafabakhsh, R.; Ha, T. Extreme bendability of DNA less than 100 base pairs long revealed by single-molecule cyclization. *Science* **2012**, *337*, 1097–1101.
19. Becker, N.A.; Kahn, J.D.; Maher, L.J. Bacterial repression loops require enhanced DNA flexibility. *J. Mol. Biol.* **2005**, *349*, 716–730.
20. Virstedt, J.; Berge, T.; Henderson, R.M.; Waring, M.J.; Travers, A.A. The influence of DNA stiffness upon nucleosome formation. *J. Struct. Biol.* **2004**, *148*, 66–85.
21. Buttinelli, M.; Minnock, A.; Panetta, G.; Waring, M.; Travers, A. The exocyclic groups of DNA modulate the affinity and positioning of the histone octamer. *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 8544–8549.
22. Diekmann, S.; von Kitzing, E.; McLaughlin, L.; Ott, J.; Eckstein, F. The influence of exocyclic substituents of purine bases on DNA curvature. *Proc. Natl. Acad. Sci. USA* **1987**, *84*, 8257–8261.
23. Nelson, H.C.; Finch, J.T.; Luisi, B.F.; Klug, A. The structure of an oligo(dA).oligo(dT) tract and its biological implications. *Nature* **1987**, *330*, 221–226.
24. Protozanova, E.; Yakovchuk, P.; Frank-Kamenetskii, M.D. Stacked-unstacked equilibrium at the nick site of DNA. *J. Mol. Biol.* **2004**, *342*, 775–785.
25. Yakovchuk, P.; Protozanova, E.; Frank-Kamenetskii, M.D. Base-stacking and base-pairing contributions into thermal stability of the DNA double helix. *Nucleic Acids Res.* **2006**, *34*, 564–574.
26. Burkhoff, A.M.; Tullius, T.D. Structural details of an adenine tract that does not cause DNA to bend. *Nature* **1988**, *331*, 455–457.
27. Haran, T.E.; Mohanty, U. The unique structure of A-tracts and intrinsic DNA bending. *Q. Rev. Biophys.* **2009**, *42*, 41–81.
28. Wiggins, P.A.; van der Heijden, T.; Moreno-Herrero, F.; Spakowitz, A.; Phillips, R.; Widom, J.; Dekker, C.; Nelson, P.C. High flexibility of DNA on short length scales probed by atomic force microscopy. *Nat. Nanotechnol.* **2006**, *1*, 137–141.



29. Savelyev, A.; Materese, C.K.; Papoian, G.A. Is DNA's rigidity dominated by electrostatic or nonelectrostatic interactions? *J. Am. Chem. Soc.* **2011**, *133*, 19290–19293.
30. Svozil, D.; Kalina, J.; Omelka, M.; Schneider, B. DNA conformations and their sequence preferences. *Nucleic Acids Res.* **2008**, *36*, 3690–3706.
31. Johnston, B.H.; Quigley, G.J.; Ellison, M.J.; Rich, A. The Z-Z junction: The boundary between two out-of-phase Z-DNA regions. *Biochemistry* **1991**, *30*, 5257–5263.
32. Wang, G.; Christensen, L.A.; Vasquez, K.M. Z-DNA-forming sequences generate large-scale deletions in mammalian cells. *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 2677–2682.
33. Liu, R.; Liu, H.; Chen, X.; Kirby, M.; Brown, P.O.; Zhao, K.J. Regulation of CSF1 promoter by the SWI/SNF-like BAF complex. *Cell* **2001**, *106*, 309–318.
34. Oh, D.B.; Kim, Y.G.; Rich, A. Z-DNA-binding proteins can act as potent effectors of gene expression *in vivo*. *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 16666–16671.
35. Garner, M.M.; Felsenfeld, G. Effect of Z-DNA on nucleosome placement. *J. Mol. Biol.* **1987**, *196*, 581–590.
36. Liu, L.F.; Wang, J.C. Supercoiling of the DNA template during transcription. *Proc. Natl. Acad. Sci. USA* **1987**, *84*, 7024–7027.
37. Herbert, A.; Lowenhaupt, K.; Spitzner, J.; Rich, A. Chicken double-stranded RNA adenosine deaminase has apparent specificity for Z-DNA. *Proc. Natl. Acad. Sci. USA* **1995**, *92*, 7550–7554.
38. Kim, Y.G.; Muralinath, M.; Brandt, T.; Percy, M.; Hauns, K.; Lowenhaupt, K.; Jacobs, B.L.; Rich, A. A role for Z-DNA binding in vaccinia virus pathogenesis. *Proc. Natl. Acad. Sci. USA* **2003**, *100*, 6974–6979.
39. Rothenburg, S.; Deigendesch, N.; Dittmar, K.; Koch-Nolte, F.; Haag, F.; Lowenhaupt, K.; Rich, A. A PKR-like eukaryotic initiation factor 2alpha kinase from zebrafish contains Z-DNA binding domains instead of dsRNA binding domains. *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 1602–1607.
40. Schwartz, T.; Behlke, J.; Lowenhaupt, K.; Heinemann, U.; Rich, A. Structure of the DLM-1-Z-DNA complex reveals a conserved family of Z-DNA-binding proteins. *Nat. Struct. Biol.* **2001**, *8*, 761–765.
41. Goodbourn, S.; Didcock, L.; Randall, R.E. Interferons: Cell signalling, immune modulation, antiviral response and virus countermeasures. *J. Gen. Virol.* **2000**, *81*, 2341–2364.
42. Peck, L.J.; Wang, J.C. Sequence dependence of the helical repeat of DNA in solution. *Nature* **1981**, *292*, 375–378.
43. Rhodes, D.; Klug, A. Sequence-dependent helical periodicity of DNA. *Nature* **1981**, *292*, 378–380.
44. Lavery, R.; Moakher, M.; Maddocks, J.H.; Petkeviciute, D.; Zakrzewska, K. Conformational analysis of nucleic acids revisited: Curves+. *Nucleic Acids Res.* **2009**, *37*, 5917–5929.
45. Struhl, K. Naturally occurring poly(dA-dT) sequences are upstream promoter elements for constitutive transcription in yeast. *Proc. Natl. Acad. Sci. USA* **1985**, *82*, 8419–8423.
46. Kaplan, N.; Moore, I.K.; Fondufe-Mittendorf, Y.; Gossett, A.J.; Tillo, D.; Field, Y.; LeProust, E.M.; Hughes, T.R.; Lieb, J.D.; Widom, J.; *et al.* The DNA-encoded nucleosome organization of a eukaryotic genome. *Nature* **2009**, *458*, 362–366.
47. Dickerson, R.E. DNA bending: The prevalence of kinkiness and the virtues of normality. *Nucleic Acids Res.* **1998**, *26*, 1906–1926.



48. Olson, W.K.; Gorin, A.A.; Lu, X.J.; Hock, L.M.; Zhurkin, V.B. DNA sequence-dependent deformability deduced from protein-DNA crystal complexes. *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 11163–11168.
49. Mack, D.R.; Chiu, T.K.; Dickerson, R.E. Intrinsic bending and deformability at the T-A step of CCTTTAAAGG: A comparative analysis of T-A and A-T steps within A-tracts. *J. Mol. Biol.* **2001**, *312*, 1037–1049.
50. Rauch, C.; Trieb, M.; Wellenzohn, B.; Loferer, M.; Voegelé, A.; Wibowo, F.R.; Liedl, K.R. C5-methylation of cytosine in B-DNA thermodynamically and kinetically stabilizes BI. *J. Am. Chem. Soc.* **2003**, *125*, 14990–14991.
51. Acosta-Silva, C.; Branchadell, V.; Bertran, J.; Oliva, A. Mutual relationship between stacking and hydrogen bonding in DNA. Theoretical study of guanine-cytosine, guanine-5-methylcytosine, and their dimers. *J. Phys. Chem. B* **2010**, *114*, 10217–10227.
52. Norberg, J.V.; Vihinen, M. Molecular dynamics simulation of the effects of cytosine methylation on structure of oligonucleotides. *J. Mol. Struct. Theochem.* **2001**, *546*, 51–62.
53. Wanunu, M.; Cohen-Karni, D.; Johnson, R.R.; Fields, L.; Benner, J.; Peterman, N.; Zheng, Y.; Klein, M.L.; Drndic, M. Discrimination of methylcytosine from hydroxymethylcytosine in DNA molecules. *J. Am. Chem. Soc.* **2011**, *133*, 486–492.
54. Hodges-Garcia, Y.; Hagerman, P.J. Cytosine methylation can induce local distortions in the structure of duplex DNA. *Biochemistry* **1992**, *31*, 7595–7599.
55. Zou, X.; Ma, W.; Solov'yov, I.A.; Chipot, C.; Schulten, K. Recognition of methylated DNA through methyl-CpG binding domain proteins. *Nucleic Acids Res.* **2012**, *40*, 2747–2758.
56. Davey, C.S.; Pennings, S.; Reilly, C.; Meehan, R.R.; Allan, J. A determining influence for CpG dinucleotides on nucleosome positioning *in vitro*. *Nucleic Acids Res.* **2004**, *32*, 4322–4331.
57. Jimenez-Useche, I.; Shim, D.; Yu, J.; Yuan, C. Unmethylated and methylated CpG dinucleotides distinctively regulate the physical properties of DNA. *Biopolymers* **2014**, *101*, 517–524.
58. Lercher, L.; McDonough, M.A.; El-Sagheer, A.H.; Thalhammer, A.; Kriaucionis, S.; Brown, T.; Schofield, C.J. Structural insights into how 5-hydroxymethylation influences transcription factor binding. *Chem. Commun.* **2014**, *50*, 1794–1796.
59. Behe, M.; Felsenfeld, G. Effects of methylation on a synthetic polynucleotide: The B-to-Z transition in poly(dG-m5dC).poly(dG-m5dC). *Proc. Natl. Acad. Sci. USA* **1981**, *78*, 1619–1623.
60. Duckett, D.R.; Murchie, A.I.; Diekmann, S.; von Kitzing, E.; Kemper, B.; Lilley, D.M.J. The structure of the Holliday junction, and its resolution. *Cell* **1988**, *55*, 79–89.
61. Liu, Y.; West, S.C. Happy Hollidays: 40th anniversary of the Holliday junction. *Nat. Rev. Mol. Cell. Biol.* **2004**, *5*, 937–944.
62. Watson, J.; Hays, F.A.; Ho, P.S. Definitions and analysis of DNA Holliday junction geometry. *Nucleic Acids Res.* **2004**, *32*, 3017–3027.
63. Mirkin, S.M. Discovery of alternative DNA structures: A heroic decade (1979–1989). *Front. Biosci.* **2008**, *13*, 1064–1071.
64. Hyeon, C.; Lee, J.; Yoon, J.; Hohng, S.; Thirumalai, D. Hidden complexity in the isomerization dynamics of Holliday junctions. *Nat. Chem.* **2012**, *4*, 907–914.
65. McKinney, S.A.; Declais, A.C.; Lilley, D.M.; Ha, T. Structural dynamics of individual Holliday junctions. *Nat. Struct. Biol.* **2003**, *10*, 93–97.

66. Gellert, M.; Lipsett, M.N.; Davies, D.R. Helix formation by guanylic acid. *Proc. Natl. Acad. Sci. USA* **1962**, *48*, 2013–2018.
67. Lam, E.Y.; Beraldi, D.; Tannahill, D.; Balasubramanian, S. G-quadruplex structures are stable and detectable in human genomic DNA. *Nat. Commun.* **2013**, *4*, 1796.
68. Todd, A.K.; Johnston, M.; Neidle, S. Highly prevalent putative quadruplex sequence motifs in human DNA. *Nucleic Acids Res.* **2005**, *33*, 2901–2907.
69. Todd, A.K.; Neidle, S. Mapping the sequences of potential guanine quadruplex motifs. *Nucleic Acids Res.* **2011**, *39*, 4917–4927.
70. Sundquist, W.I.; Klug, A. Telomeric DNA dimerizes by formation of guanine tetrads between hairpin loops. *Nature* **1989**, *342*, 825–829.
71. Biffi, G.; Tannahill, D.; McCafferty, J.; Balasubramanian, S. Quantitative visualization of DNA G-quadruplex structures in human cells. *Nat. Chem.* **2013**, *5*, 182–186.
72. Koole, W.; van Schendel, R.; Karambelas, A.E.; van Heteren, J.T.; Okihara, K.L.; Tijsterman, M.A. A Polymerase Theta-dependent repair pathway suppresses extensive genomic instability at endogenous G4 DNA sites. *Nat. Commun.* **2014**, *5*, 3216.
73. Zhao, J.; Bacolla, A.; Wang, G.; Vasquez, K.M. Non-B DNA structure-induced genetic instability and evolution. *Cell. Mol. Life Sci.* **2010**, *67*, 43–62.
74. London, T.B.; Barber, L.J.; Mosedale, G.; Kelly, G.P.; Balasubramanian, S.; Hickson, I.D.; Boulton, S.J.; Hiom, K. FANCD1 is a structure-specific DNA helicase associated with the maintenance of genomic G/C tracts. *J. Biol. Chem.* **2008**, *283*, 36132–36139.
75. Popuri, V.; Bachrati, C.Z.; Muzzolini, L.; Mosedale, G.; Costantini, S.; Giacomini, E.; Hickson, I.D.; Vindigni, A. The Human RecQ helicases, BLM and RECQ1, display distinct DNA substrate specificities. *J. Biol. Chem.* **2008**, *283*, 17766–17776.
76. Eddy, S.; Ketkar, A.; Zafar, M.K.; Maddukuri, L.; Choi, J.Y.; Eoff, R.L. Human Rev1 polymerase disrupts G-quadruplex DNA. *Nucleic Acids Res.* **2014**, *42*, 3272–3285.
77. Lopes, J.; Piazza, A.; Bermejo, R.; Kriegsman, B.; Colosio, A.; Teulade-Fichou, M.P.; Foiani, M.; Nicolas, A. G-quadruplex-induced instability during leading-strand replication. *EMBO J.* **2011**, *30*, 4033–4046.
78. Ribeyre, C.; Lopes, J.; Boule, J.B.; Piazza, A.; Guedin, A.; Zakian, V.A.; Mergny, J.L.; Nicolas, A. The yeast Pif1 helicase prevents genomic instability caused by G-quadruplex-forming CEB1 sequences *in vivo*. *PLoS Genet.* **2009**, *5*, e1000475.
79. Jen-Jacobson, L. Protein-DNA recognition complexes: Conservation of structure and binding energy in the transition state. *Biopolymers* **1997**, *44*, 153–180.
80. Lin, S.; Riggs, A.D. A comparison of lac repressor binding to operator and to nonoperator DNA. *Biochem. Biophys. Res. Commun.* **1975**, *62*, 704–710.
81. von Hippel, P.H.; Berg, O.G. On the specificity of DNA-protein interactions. *Proc. Natl. Acad. Sci. USA* **1986**, *83*, 1608–1612.
82. Härd, T.; Lundbäck, T. Thermodynamics of sequence-specific protein-DNA interactions. *Biophys. Chem.* **1996**, *62*, 121–139.
83. Jen-Jacobson, L.; Engler, L.E.; Jacobson, L.A. Structural and thermodynamic strategies for site-specific DNA binding proteins. *Structure* **2000**, *8*, 1015–1023.

84. Travers, A.A.; Vaillant, C.; Arneodo, A.; Muskhelishvili, G. DNA structure, nucleosome placement and chromatin remodelling: A perspective. *Biochem. Soc. Trans.* **2012**, *40*, 335–340.
85. Rohs, R.; Jin, X.; West, S.M.; Joshi, R.; Honig, B.; Mann, R.S. Origins of specificity in protein-DNA recognition. *Annu. Rev. Biochem.* **2010**, *79*, 233–269.
86. Lavery, R. Recognizing DNA. *Q. Rev. Biophys.* **2005**, *38*, 339–344.
87. Seeman, N.C.; Rosenberg, J.M.; Rich, A. Sequence-specific recognition of double helical nucleic acids by proteins. *Proc. Natl. Acad. Sci. USA* **1976**, *73*, 804–808.
88. Coulocheri, S.A.; Pigis, D.G.; Papavassiliou, K.A.; Papavassiliou, A.G. Hydrogen bonds in protein-DNA complexes: Where geometry meets plasticity. *Biochimie* **2007**, *89*, 1291–1303.
89. Gordan, R.; Shen, N.; Dror, I.; Zhou, T.; Horton, J.; Rohs, R.; Bulyk, M.L. Genomic regions flanking E-box binding sites influence DNA binding specificity of bHLH transcription factors through DNA shape. *Cell. Rep.* **2013**, *3*, 1093–1104.
90. Harris, L.A.; Watkins, D.; Williams, L.D.; Koudelka, G.B. Indirect readout of DNA sequence by p22 repressor: Roles of DNA and protein functional groups in modulating DNA conformation. *J. Mol. Biol.* **2013**, *425*, 133–143.
91. Shakked, Z.; Guzikevich-Guerstein, G.; Frolov, F.; Rabinovich, D.; Joachimiak, A.; Sigler, P.B. Determinants of repressor/operator recognition from the structure of the trp operator binding site. *Nature* **1994**, *368*, 469–473.
92. Romanuka, J.; Folkers, G.E.; Biris, N.; Tishchenko, E.; Wienk, H.; Bonvin, A.M.; Kaptein, R.; Boelens, R. Specificity and affinity of Lac repressor for the auxiliary operators O2 and O3 are explained by the structures of their protein-DNA complexes. *J. Mol. Biol.* **2009**, *390*, 478–489.
93. Loenen, W.A.; Dryden, D.T.; Raleigh, E.A.; Wilson, G.G.; Murray, N.E. Highlights of the DNA cutters: A short history of the restriction enzymes. *Nucleic Acids Res.* **2014**, *42*, 3–19.
94. Williams, R.J. Restriction endonucleases: Classification, properties, and applications. *Mol. Biotechnol.* **2003**, *23*, 225–243.
95. Pingoud, A.; Jeltsch, A. Structure and function of type II restriction endonucleases. *Nucleic Acids Res.* **2001**, *29*, 3705–3727.
96. Watanabe, N.; Takasaki, Y.; Sato, C.; Ando, S.; Tanaka, I. Structures of restriction endonuclease *HindIII* in complex with its cognate DNA and divalent cations. *Acta Crystallogr. D Biol. Crystallogr.* **2009**, *65*, 1326–1333.
97. Kostrewa, D.; Winkler, F.K. Mg<sup>2+</sup> binding to the active site of *EcoRV* endonuclease: A crystallographic study of complexes with substrate and product DNA at 2 Å resolution. *Biochemistry* **1995**, *34*, 683–696.
98. Roszczyk, E.; Goodgal, S. Methylase activities from *Haemophilus influenzae* that protect *Haemophilus parainfluenzae* transforming deoxyribonucleic acid from inactivation by *Haemophilus influenzae* endonuclease R. *J. Bacteriol.* **1975**, *123*, 287–293.
99. Nwankwo, D.O.; Moran, L.S.; Slatko, B.E.; Waite-Rees, P.A.; Dorner, L.F.; Benner, J.S.; Wilson, G.G. Cloning, analysis and expression of the *HindIII* R-M-encoding genes. *Gene* **1994**, *150*, 75–80.
100. Martin, A.M.; Sam, M.D.; Reich, N.O.; Perona, J.J. Structural and energetic origins of indirect readout in site-specific DNA cleavage by a restriction endonuclease. *Nat. Struct. Biol.* **1999**, *6*, 269–277.

101. Martin, A.M.; Horton, N.C.; Lusetti, S.; Reich, N.O.; Perona, J.J. Divalent metal dependence of site-specific DNA binding by *EcoRV* endonuclease. *Biochemistry* **1999**, *38*, 8430–8439.
102. Otwinowski, Z.; Schevitz, R.W.; Zhang, R.G.; Lawson, C.L.; Joachimiak, A.; Marmorstein, R.Q.; Luisi, B.F.; Sigler, P.B. Crystal structure of trp repressor/operator complex at atomic resolution. *Nature* **1988**, *335*, 321–329.
103. Smith, S.A.; Rajur, S.B.; McLaughlin, L.W. Specific purine N7-nitrogens are critical for high affinity binding by the Trp repressor. *Nat. Struct. Biol.* **1994**, *1*, 198.
104. Zhang, A.P.; Pigli, Y.Z.; Rice, P.A. Structure of the LexA-DNA complex and implications for SOS box measurement. *Nature* **2010**, *466*, 883–886.
105. Fernandez De Henestrosa, A.R.; Ogi, T.; Aoyagi, S.; Chafin, D.; Hayes, J.J.; Ohmori, H.; Woodgate, R. Identification of additional genes belonging to the LexA regulon in *Escherichia coli*. *Mol. Microbiol.* **2000**, *35*, 1560–1572.
106. Wade, J.T.; Reppas, N.B.; Church, G.M.; Struhl, K. Genomic analysis of LexA binding reveals the permissive nature of the *Escherichia coli* genome and identifies unconventional target sites. *Genes Dev.* **2005**, *19*, 2619–2630.
107. White, R.J.; Jackson, S.P. The TATA-binding protein: A central role in transcription by RNA polymerases I, II and III. *Trends Genet.* **1992**, *8*, 284–288.
108. Sharp, P.A. TATA-binding protein is a classless factor. *Cell* **1992**, *68*, 819–821.
109. Rigby, P.W. Three in one and one in three: It all depends on TBP. *Cell* **1993**, *72*, 7–10.
110. Comai, L.; Tanese, N.; Tjian, R. The TATA-binding protein and associated factors are integral components of the RNA polymerase I transcription factor, SL1. *Cell* **1992**, *68*, 965–976.
111. Kim, J.L.; Burley, S.K. 1.9 Å resolution refined structure of TBP recognizing the minor groove of TATAAAAG. *Nat. Struct. Biol.* **1994**, *1*, 638–653.
112. Kim, Y.; Geiger, J.H.; Hahn, S.; Sigler, P.B. Crystal structure of a yeast TBP/TATA-box complex. *Nature* **1993**, *365*, 512–520.
113. Arfin, S.M.; Long, A.D.; Ito, E.T.; Toller, L.; Riehle, M.M.; Paegle, E.S.; Hatfield, G.W. Global gene expression profiling in *Escherichia coli* K12. The effects of integration host factor. *J. Biol. Chem.* **2000**, *275*, 29672–29684.
114. Kobryn, K.; Lavoie, B.D.; Chaconas, G. Supercoiling-dependent site-specific binding of HU to naked Mu DNA. *J. Mol. Biol.* **1999**, *289*, 777–784.
115. Hwang, D.S.; Kornberg, A. Opening of the replication origin of *Escherichia coli* by DnaA protein with protein HU or IHF. *J. Biol. Chem.* **1992**, *267*, 23083–23086.
116. Kur, J.; Hasan, N.; Szybalski, W. Physical and biological consequences of interactions between integration host factor (IHF) and coliphage lambda late p'R promoter and its mutants. *Gene* **1989**, *81*, 1–15.
117. Hillyard, D.R.; Edlund, M.; Hughes, K.T.; Marsh, M.; Higgins, N.P. Subunit-specific phenotypes of *Salmonella typhimurium* HU mutants. *J. Bacteriol.* **1990**, *172*, 5402–5407.
118. Li, S.; Waters, R. *Escherichia coli* strains lacking protein HU are UV sensitive due to a role for HU in homologous recombination. *J. Bacteriol.* **1998**, *180*, 3750–3756.
119. Miyabe, I.; Zhang, Q.M.; Kano, Y.; Yonei, S. Histone-like protein HU is required for recA gene-dependent DNA repair and SOS induction pathways in UV-irradiated *Escherichia coli*. *Int. J. Radiat. Biol.* **2000**, *76*, 43–49.

120. Johnson, R.C.; Johnson, L.M.; Schmidt, J.; Garder, J.F. The major nucleoid proteins in the structure and function of the *Escherichia coli* chromosome. In *The Bacterial Chromosomes*; ASM Press: Washington, DC, USA, 2005; Volume 1, pp. 65–132.
121. Pil, P.M.; Chow, C.S.; Lippard, S.J. High-mobility-group 1 protein mediates DNA bending as determined by ring closures. *Proc. Natl. Acad. Sci. USA* **1993**, *90*, 9465–9469.
122. Paull, T.T.; Haykinson, M.J.; Johnson, R.C. The nonspecific DNA-binding and -bending proteins HMG1 and HMG2 promote the assembly of complex nucleoprotein structures. *Genes Dev.* **1993**, *7*, 1521–1534.
123. Shimizu, M.; Miyake, M.; Kanke, F.; Matsumoto, U.; Shindo, H. Characterization of the binding of HU and IHF, homologous histone-like proteins of *Escherichia coli*, to curved and uncurved DNA. *Biochim. Biophys. Acta* **1995**, *1264*, 330–336.
124. Swinger, K.K.; Rice, P.A. Structure-based analysis of HU-DNA binding. *J. Mol. Biol.* **2007**, *365*, 1005–1016.
125. Lynch, T.W.; Read, E.K.; Mattis, A.N.; Gardner, J.F.; Rice, P.A. Integration host factor: Putting a twist on protein-DNA recognition. *J. Mol. Biol.* **2003**, *330*, 493–502.
126. Dhavan, G.M.; Crothers, D.M.; Chance, M.R.; Brenowitz, M. Concerted binding and bending of DNA by *Escherichia coli* integration host factor. *J. Mol. Biol.* **2002**, *315*, 1027–1037.
127. Bosch, D.; Campillo, M.; Pardo, L. Binding of proteins to the minor groove of DNA: What are the structural and energetic determinants for kinking a basepair step? *J. Comput. Chem.* **2003**, *24*, 682–691.
128. Grove, A.; Saavedra, T.C. The role of surface-exposed lysines in wrapping DNA about the bacterial histone-like protein HU. *Biochemistry* **2002**, *41*, 7597–7603.
129. Holbrook, J.A.; Tsodikov, O.V.; Saecker, R.M.; Record, M.T. Specific and non-specific interactions of integration host factor with DNA: Thermodynamic evidence for disruption of multiple IHF surface salt-bridges coupled to DNA binding. *J. Mol. Biol.* **2001**, *310*, 379–401.
130. Lowary, P.T.; Widom, J. New DNA sequence rules for high affinity binding to histone octamer and sequence-directed nucleosome positioning. *J. Mol. Biol.* **1998**, *276*, 19–42.
131. Satchwell, S.C.; Drew, H.R.; Travers, A.A. Sequence periodicities in chicken nucleosome core DNA. *J. Mol. Biol.* **1986**, *191*, 659–675.
132. Dayn, A.; Malkhosyan, S.; Mirkin, S.M. Transcriptionally driven cruciform formation *in vivo*. *Nucleic Acids Res.* **1992**, *20*, 5991–5997.
133. Lobachev, K.S.; Rattray, A.; Narayanan, V. Hairpin- and cruciform-mediated chromosome breakage: Causes and consequences in eukaryotic cells. *Front. Biosci.* **2007**, *12*, 4208–4220.
134. Maher, R.L.; Branagan, A.M.; Morriscal, S.W. Coordination of DNA replication and recombination activities in the maintenance of genome stability. *J. Cell. Biochem.* **2011**, *112*, 2672–2682.
135. Xu, J.; de Zhu, J.; Ni, M.; Wan, F.; Gu, J.R. The ATF/CREB site is the key element for transcription of the human RNA methyltransferase like 1(RNMTL1) gene, a newly discovered 17p13.3 gene. *Cell Res.* **2002**, *12*, 177–197.
136. Hanke, J.H.; Hambor, J.E.; Kavathas, P. Repetitive Alu elements form a cruciform structure that regulates the function of the human CD8 alpha T cell-specific enhancer. *J. Mol. Biol.* **1995**, *246*, 63–73.

137. Van Brabant, A.J.; Stan, R.; Ellis, N.A. DNA helicases, genomic instability, and human genetic disease. *Annu. Rev. Genomics Hum. Genet.* **2000**, *1*, 409–459.
138. Alvarez, D.; Novac, O.; Callejo, M.; Ruiz, M.T.; Price, G.B.; Zannis-Hadjopoulos, M. 14-3-3sigma is a cruciform DNA binding protein and associates *in vivo* with origins of DNA replication. *J. Cell. Biochem.* **2002**, *87*, 194–207.
139. Callejo, M.; Alvarez, D.; Price, G.B.; Zannis-Hadjopoulos, M. The 14-3-3 protein homologues from *Saccharomyces cerevisiae*, Bmh1p and Bmh2p, have cruciform DNA-binding activity and associate *in vivo* with ARS307. *J. Biol. Chem.* **2002**, *277*, 38416–38423.
140. Zannis-Hadjopoulos, M.; Yahyaoui, W.; Callejo, M. 14-3-3 cruciform-binding proteins as regulators of eukaryotic DNA replication. *Trends Biochem. Sci.* **2008**, *33*, 44–50.
141. Wakasugi, M.; Reardon, J.T.; Sancar, A. The non-catalytic function of XPG protein during dual incision in human nucleotide excision repair. *J. Biol. Chem.* **1997**, *272*, 16030–16034.
142. Lee, J.H.; Park, C.J.; Arunkumar, A.I.; Chazin, W.J.; Choi, B.S. NMR study on the interaction between RPA and DNA decamer containing cis-syn cyclobutane pyrimidine dimer in the presence of XPA: Implication for damage verification and strand-specific dual incision in nucleotide excision repair. *Nucleic Acids Res.* **2003**, *31*, 4747–4754.
143. Sekelsky, J.J.; Hollis, K.J.; Eimerl, A.I.; Burtis, K.C.; Hawley, R.S. Nucleotide excision repair endonuclease genes in *Drosophila melanogaster*. *Mutat. Res.* **2000**, *459*, 219–228.
144. Stefanovsky, V.Y.; Moss, T. The cruciform DNA mobility shift assay: A tool to study proteins that recognize bent DNA. *Methods Mol. Biol.* **2009**, *543*, 537–546.
145. Van Houte, L.P.; Chuprina, V.P.; van der Wetering, M.; Boelens, R.; Kaptein, R.; Clevers, H. Solution structure of the sequence-specific HMG box of the lymphocyte transcriptional activator Sox-4. *J. Biol. Chem.* **1995**, *270*, 30516–30524.
146. Pearson, C.E.; Ruiz, M.T.; Price, G.B.; Zannis-Hadjopoulos, M. Cruciform DNA binding protein in HeLa cell extracts. *Biochemistry* **1994**, *33*, 14185–14196.
147. Declais, A.C.; Lilley, D.M. New insight into the recognition of branched DNA structure by junction-resolving enzymes. *Curr. Opin. Struct. Biol.* **2008**, *18*, 86–95.
148. Lilley, D.M.; White, M.F. The junction-resolving enzymes. *Nat. Rev. Mol. Cell. Biol.* **2001**, *2*, 433–443.
149. Khuu, P.A.; Voth, A.R.; Hays, F.A.; Ho, P.S. The stacked-X DNA Holliday junction and protein recognition. *J. Mol. Recognit.* **2006**, *19*, 234–242.
150. Lilley, D.M. Structures of helical junctions in nucleic acids. *Q. Rev. Biophys.* **2000**, *33*, 109–159.
151. Brill, S.J. Linking the enzymes that unlink DNA. *Mol. Cell.* **2013**, *52*, 159–160.
152. Brazda, V.; Laister, R.C.; Jagelska, E.B.; Arrowsmith, C. Cruciform structures are a common DNA feature important for regulating biological processes. *BMC Mol. Biol.* **2011**, *12*, 1–16.
153. Gorecka, K.M.; Komorowska, W.; Nowotny, M. Crystal structure of RuvC resolvase in complex with Holliday junction substrate. *Nucleic Acids Res.* **2013**, *41*, 9945–9955.
154. Biertumpfel, C.; Yang, W.; Suck, D. Crystal structure of T4 endonuclease VII resolving a Holliday junction. *Nature* **2007**, *449*, 616–620.
155. Ortiz-Lombardia, M.; Gonzalez, A.; Eritja, R.; Aymami, J.; Azorin, F.; Coll, M. Crystal structure of a DNA Holliday junction. *Nat. Struct. Biol.* **1999**, *6*, 913–917.

156. Mashal, R.D.; Koontz, J.; Sklar, J. Detection of mutations by cleavage of DNA heteroduplexes with bacteriophage resolvases. *Nat. Genet.* **1995**, *9*, 177–183.
157. Hadden, J.M.; Declais, A.C.; Carr, S.B.; Lilley, D.M.; Phillips, S.E. The structural basis of Holliday junction resolution by T7 endonuclease I. *Nature* **2007**, *449*, 621–624.
158. Declais, A.C.; Liu, J.; Freeman, A.D.; Lilley, D.M. Structural recognition between a four-way DNA junction and a resolving enzyme. *J. Mol. Biol.* **2006**, *359*, 1261–1276.
159. Freeman, A.D.; Declais, A.C.; Lilley, D.M. The importance of the *N*-terminus of T7 endonuclease I in the interaction with DNA junctions. *J. Mol. Biol.* **2013**, *425*, 395–410.
160. Herbert, A.; Alfken, J.; Kim, Y.G.; Mian, I.S.; Nishikura, K.; Rich, A.A. A Z-DNA binding domain present in the human editing enzyme, double-stranded RNA adenosine deaminase. *Proc. Natl. Acad. Sci. USA* **1997**, *94*, 8421–8426.
161. Herbert, A.G.; Spitzner, J.R.; Lowenhaupt, K.; Rich, A. Z-DNA binding protein from chicken blood nuclei. *Proc. Natl. Acad. Sci. USA* **1993**, *90*, 3339–3342.
162. Samuel, C.E. Adenosine deaminases acting on RNA (ADARs) are both antiviral and proviral. *Virology* **2011**, *411*, 180–193.
163. Schwartz, T.; Rould, M.A.; Lowenhaupt, K.; Herbert, A.; Rich, A. Crystal structure of the Zalpha domain of the human editing enzyme ADAR1 bound to left-handed Z-DNA. *Science* **1999**, *284*, 1841–1845.
164. De Rosa, M.; de Sanctis, D.; Rosario, A.L.; Archer, M.; Rich, A.; Athanasiadis, A.; Carrondo, M.A. Crystal structure of a junction between two Z-DNA helices. *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 9088–9092.