Hoogsteen Base Pairings: A New Paradigm for DNA Replication, DNA Recognition, and DNA Repair

Neelabh Datta*

Department of Biochemistry, Asutosh College (Affiliated to University of Calcutta), 92, Shyamaprasad Mukherjee Road, Kolkata 700026, West Bengal, India
Corresponding Author: Neelabh Datta, *neelabhdatta@gmail.com

Abstract

In contrast to Watson-Crick (WC) base pairing, Hoogsteen (HG) base pairing involves flipping a purine base 180° between its anti and syn conformation. Recent studies have shown that HG pairs coexist in dynamical equilibrium, and several biological functions depend on them. This significance has stirred computational research on this base-pairing transition. However, a methodical reproduction of sequence variations has continued to be out of reach. It is becoming increasingly clear that Hoogsteen base pairs play a crucial role in DNA replication, recognition, damage repair, and incorrect sequence repair. The Protein Data Bank contains a variety of Hoogsteen base pairing modes that include the preference for A–T versus G–C bps, TA versus GG pairs, and a preference for 5'-purines at terminal ends. RNA A-form duplexes are strongly disfavoured by Hoogsteen base pairs, in stark contrast to B-form DNA. Therefore, N1-methyl adenosine and N1-methyl guanosine, which is found in DNA as alkylation impairment and in RNA as posttranscriptional adjustments, have great differences in effects. They create G–C+ and A–U Hoogsteen base pairs in duplex DNA that preserve the structural integrity of the double helix but obstruct base pairing altogether and induce local duplex melting in RNA, providing a mechanism for potently disrupting RNA structure through posttranscriptional modifications. In duplex DNA, they maintain the structural integrity of the double helix by creating G–C+ and A–U Hoogsteen base pairs, but block base pairing altogether and cause local duplex melting in RNA, thus providing a potent means for disrupting RNA structure post transcriptionally. As a result of the markedly different inclinations for B-DNA and A-RNA to form Hoogsteen base pairs, they may be able to balance the opposing demands of maintaining genome stability and dynamically modulating the epitranscriptome. This review examines the occurrence of Hoogsteen base pairs in DNA and RNA duplexes.

Keywords: Hoogsteen bp, DNA structure, DNA dynamics, DNA sequencing, Hoogsteen bp dynamics

Introduction

In 1953, characteristic studies by Watson & Crick, 1953 demarcated the arrangement of DNA for the first time. With the discovery of hydrogen bonds between purine and pyrimidine bases, DNA's double-helical shape was revealed, along with its replication process as an essential part of its function as a biological information carrier. Grounded on crystallographic data from A-T crystals, Hoogsteen projected an unconventional base-pairing motif in 1959, in which the purine base rolls 180° around the glycosidic bond (Hoogsteen,1959), i.e. from anti to syn, with respect to the Watson-Crick-Franklin WCF geometry (Ortíz et al., 2022).

The pyrimidine in the Hoogsteen (HG) configuration forms hydrogen bonds with the purine's 5-ring rather than its 6-ring, resulting in a shorter distance between the opposite C1' atoms of the bases as well as some bending and twisting of the double helix around the base pair (Hoogsteen, 1959, Zhou et al., 2015). Several studies have shown in the last few years that canonical duplex DNA frequently possesses the HG conformation and that its biological significance is considerable. By using nuclear magnetic resonance relaxation dispersion spectroscopy (NMR RDS) in 2011, Nikolova and co-workers detected transient HG base pairs inside canonical duplex DNA (Nikolova et al., 2011). They conveyed populations of A·T and G·C HG base pairs of about 0.5%, with residence intervals of up to 1.5 ms (Ortíz et al., 2022). Modern research has measured even superior HG populations, of 1.2%, in an A·T rich segment (Imeddourene et al., 2020). Alvey and colleagues validated that HG base pairs occur in more assorted sequences than formerly assumed using NMR relaxation dispersion a few years later (Alvey et al., 2014).

HG base pairs, which had originally been thought to occur only in damaged DNA, are now considered to coexist in dynamic equilibrium with the WCF bases in recent studies (Zhou et al., 2015, Nikolova et al., 2013). Some of these complexes are of precise genetic significance (Ortíz et al., 2022). As an example, the human DNA polymerase-ι can perform replication exclusively using HG base pairing (Nair et al., 2004, Johnson et al,, 2005). It has also been shown that HG functions in DNA-protein complexes associated with p53 tumour suppressor protein, TATA-box binding protein involved in transcription, and the MATα2 homeo-domain which regulates transcription in cells (Kitayner et al., 2010, Patikoglou et al., 1999, Aishima et al., 2002). HG base pairs have recently been discovered to be relevant, and their coexistence with WCF base pairs requires a deeper

Received 2022-11-10; revised 2022-11-28; accepted 2022-11-28; published online 2022-12-08; published 2022-12-23

understanding of the transition mechanisms between both forms. In the absence of experimental observation of short-lived intermediate states, computational approaches using molecular dynamics simulations, boosted with enhanced sampling, have proved useful.

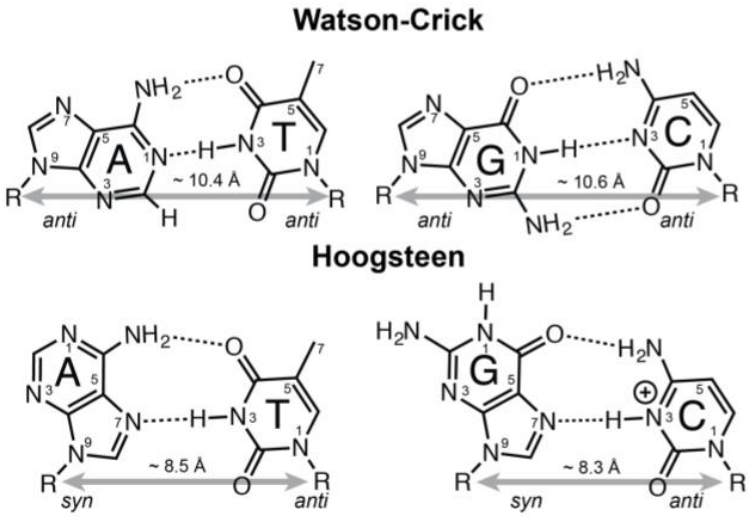

Figure 1. Base Pairings of WC bp and HG bpSource: Zhou, Huiqing (2016). Occurrence and Function of Hoogsteen Base Pairs in Nucleic Acids. Dissertation, Duke University.

Nikolova and co-workers supplemented their NMR results in 2011 with Conjugate Peak Refinement simulations (CPR) (Ortíz et al., 2022). They used the CHARMM27 force field to simulate the DNA sequence 5'-CGATTTTTTGGGC-3' (A6-DNA) in vacuo and studied the transition between the A16 T9 pairs (MacKerell et al., 2000). For steering the transition, two collective variables (CVs) were used: (1) χ glycosidic angle, which designates the rolling of A16 around the glycosidic bond and is demarcated by the atoms O4'-C1'-N9-C4; (2) θ base opening angle, which describes the flipping of A16 from the double helix toward the major groove and is defined in (Son et al., 2009). A CPR simulation revealed two types of pathways: one in which the adenine rolls clockwise or counterclockwise inside the double helix, with a small opening angle at the base; the other is outside the double helix, with a large opening angle at the base. These two types of pathways are referred to as inside and outside, respectively. In their (χ, θ) pseudo-free-energy landscapes, Nikolova and co-workers seemed to prefer the inside mechanism more than the outside mechanism, which is probably due to the lack of solvent to stabilize the flipped conformations (Ortíz et al., 2022). An umbrella sampling (US) study was conducted in 2015 by Yang and colleagues to determine the (χ, θ) free-energy landscapes of A16 and T9 for the same A-DNA pair (Yang et al., 2015). Their analysis used an AMBER99-BSC0 force field with modified glycosidic torsion parameters, as well as explicit TIP3P water (Perez et al., 2007). With over 300 windows, their US calculations revealed multiple pathways, including inside and outside routes that lead to either the major or minor grooves (Ortíz et al., 2022). Yang and co-workers reported a free-energy difference from the WCF to the HG state of 4.4 kcal/mol, close to Nikolova and co-workers' NMR result of 3.0 kcal/mol. Similarly, Ray et al., 2020 calculated a (χ, θ) free-energy surface for DNA and RNA. According to them, DNA and RNA surface (χ, θ) free-energy barrier is of the order of 10-11 kcal/mol in the inside pathways and can reach 14 kcal/mol in the outside pathways. The systematic study of multiple DNA variations in 2D free-energy landscapes is hindered by the significant computational expense of obtaining several μs-long MD runs. There is experimental confirmation that specific sequence patterns, such as AT steps, can favour HG base pairing. However, it has remained past computational reach to study the effect of the DNA sequence on base pairing proclivity (Acosta-Reyes et al., 2015).

Hoogsteen pairings in naked duplexes

The early 1980s saw the proposal of HG bps as the sole component of Z-DNA27, especially for AT-rich sequences that exhibited irregular X-ray diffraction patterns when dried (referred to as D- or E-type diffraction patterns) (Davies & Baldwin, 1963, Drew & Dickerson, 1982). This form of DNA entails helical structures with 7–7.5 bps per turn, which cannot be chemically attained by right-handed B-form DNA (Zhou, 2016). According to spectroscopic studies of poly(rA)-poly(rU) sequences accept substituents at adenine C2 that block WC base pairing, duplexes with parallel or antiparallel chain polarity are also formed, with HG or reverse HG base pairings holding the strands together (Ikehara et al., 2005, Ishikawa et al., 1972). During the 1990s, the development of $^{13}C/^{15}N$ isotopic enrichment and solution-state NMR spectroscopy of nucleic acids, coupled with the ease of preparing large quantities of highly pure DNA samples, enabled high-resolution X-ray and NMR structure determination of diverse DNA sequences showing WC B-form DNA duplexes. As a result, spectroscopic

evidence for HG bps continued to emerge in the 1990s and 2000s in A-T rich sequences (Liu et al., 1993), poly(dG-dC)-poly(dG-dC) sequences at low pH as potential intermediates alongside the B-to-Z DNA transition (Segers-Nolten et al., 1997), as well as non-canonical DNA regions such as closing bps of apical loops (Blommers et al., 1991, Ronning et al., 2005). The structure of $d(AT)_3$ exposed an anti-parallel right-handed double helix made up entirely of HG bps with a complete structure analogous to that of B-form DNA (Abrescia et al., 2002, Abrescia et al., 2004). This variation was characterized by a shift in the helical axis relative to the bps, a reduction in the helical radius and C1-'C1' by ≈2.5–3.0 Å, an altered hydrogen bonding donor/acceptor pattern in both major and minor grooves, and a narrower, less electronegative minor groove that favours hydrophobic contacts and distinct helix stacking and hydration patterns relative to B-DNA (Zhou, 2016). In conjunction, these features provide a distinct physicochemical presentation of the genetic code that can potentially be recognized by cellular machinery based on sequence-specific characteristics. Nevertheless, no naked AT-repeats in DNA oligonucleotides have crystallized as WC helixes; however, solution-state NMR studies of the above DNA sequences in aqueous solution or under the same conditions used to grow crystals argued against the HG helix formation (Abrescia et al., 2004). According to the solution NMR studies, AT-repeat DNA duplexes possess prototypical WC B-form double helices while they are highly unstable in solution with melting temperatures below 5°C (Zhou, 2016). This shows that despite crystal packing possibly playing a key role in stabilizing the HG double helix, it cannot be excluded that WC to HG transitions may occur more frequently in AT-repeat sequences than in other sequences.

Hoogsteen pairings in DNA-Protein complexes

In the late 1990s, X-ray structures were developed which depicted that certain proteins bind and, in some circumstances, explicitly recognize HG bps embedded in B-form DNA. A functional role for HG bps in vivo has been demonstrated by these studies, which suggest that proteins exploit the unique structural and chemical properties of HG bps in sequence-specific DNA recognition. The X-ray structure of a highly twisted (>160 degrees) 35 bp DNA attached to the integration host factor (IHF) protein exhibited a single A-T HG bp adjacent to a nicked site in 1996 (Rice et al., 1996). In the HG basis pair, the WC face was specifically recognized by a hydrogen bond between the amide group of the backbone of the arginine residue and the third hydrogen group of the syn A. As a result of HG formation, the peptide backbone is moved away from a neighbouring molecule in the complex by the nick, which also aids in crystal packing (Zhou, 2016). The protein also makes explicit contacts with N3 of an anti-A in a symmetric site in the DNA without a nick, suggesting interactions specific to WC rather than HG. In subsequent X-ray structures of TATA elements bound to TATA box-binding protein (TBP), a G–C+ HG bp was observed in the mutant TATAAAC box in a region of DNA unwinding and intercalation (Rangadurai et al., 2022). However direct interactions between the syn guanine base and TBP were not observed. While preserving Van der Waals's contacts with two neighbouring phenylalanine residues, the HG bp appears to contribute to binding by preventing steric clashes between leucine 72 and the exocyclic guanine NH2 (Zhou, 2016).

A second G–C+ HG bp was detected but accredited to crystal packing forces. Remarkably, the ≈150-fold weaker binding affinity observed for TBP to this mutant TATA box (Hoopes et al., 1998) which could be associated with the selection of a transient HG over a WC bp at that site (Ol'ha et al., 2021) has been concerned in the transcriptional regulation of the human osteocalcin gene (Meyer et al., 1997). This observation advocates a biological role in the formation of a $G–C^+$ HG bp at the mutant promoter site. Both IHF and TBP together bring about large distortions in the DNA, which could assist the development of HG bps. HG bps are accommodated within duplex DNA without causing major distortions, even for directly neighbouring bps, avoiding unfavourable steric clashes associated with WC bps. In view of the ease with which HG bps can be seamlessly incorporated into B-DNA, HG bps might have been misinterpreted as WC bps by misinterpreting ambiguous electron densities at low to medium resolution (Robert et al., 2021). Two neighbouring A-T HG bps were consequently perceived in structures of a palindromic CATG/CATG sequence attached to the DNA binding domain of p53 (Vreede et al., 2019). HG bps form a narrowed minor groove flanking the CATG site that leads to enhanced negative electrostatic potential, which is further stabilized by the addition of positively charged arginine side chains, despite no direct contacts being observed with the syn adenines. X-ray structures show that these HG bps adopt WC conformation with longer spacers (Vreede et al., 2019) or altered superseding sequences (Malecka et al., 2009, Chen et al., 2010) between DNA half-sites. This is complemented by altered DNA helix conformation, altered DNA dimer organization, and altered DNA-tetramer binding affinity. As a result of these studies, WC and HG bps likely exist in equilibrium with each other, and their selection in DNA-p53 complexes is largely determined by the sequence of DNA bound to the p53 protein.

Hoogsteen pairings in DNA replication

In Watson and Crick's 1953 Nature paper, Watson and Crick express their belief that WC bps, the most significant part of the DNA double helix structure, indicates a possibility of copying genetic material. The enzyme that catalyses template DNA replication was discovered four years later by Lehman (Lehman et al., 1958) and subsequent studies demonstrated that DNA polymerases replicate DNA by pairing dNTP with template strands by WC pairing. During the 1990s, studies revealed that certain families of DNA polymerases like the X and Y families subsidised damage-induced mutagenesis. It was revealed

later that the Y family members resourcefully evade DNA damage by replicating template DNA using HG rather than WC base pairing. More than a decade ago, Yang and colleagues first observed HG-based replication in X-ray structures of a DNA polymerase homolog DNA Polι, Dpo4 (Ling et al., 2003). By forming a HG bp between the 5' thymine and incoming dATP, Dpo4 replicates UV cross-linked thymine dimers, avoiding backbone distortion, and discriminating against guanine and pyrimidine (Ling et al., 2003). Nair and co-workers successively presented using X-ray crystallography and biochemical experimentations that alternative member of this family, DNA Polι, employs HG base-paring as a common apparatus to replicate both damaged and undamaged DNA (Nair et al., 2004). A prior biochemical study demonstrating a much higher efficiency of correct base incorporation across a templating adenine than across templating thymine also provided a rationale for inserting the correct nucleotide across an adenine base because it is highly like to form anti-G–T wobble bps, it favours G disincorporation (Tissier et al., 2000). This raised HG bps to a prominent point set aside formerly only for WC bps; they provided a basis for copying DNA. The proposal that Polι replicates DNA via HG base-pairing was quickly met with scepticism. In a News & View article, Wang, 2005 pointed out that since the A–T base pairs are the weakest electron densities at active sites, it is difficult to distinguish between WC and HG conformations. In addition to the original structure, several other structures were derived that exposed that major purine alkylation and oxidation lesions, including 1,N6-ethenoadenine, N2 ethylguanine, O6-methylguanine, and 8 oxoguanine (Zhou, 2016), adopted syn conformation and, where conceivable, formed HG type bps with complementary purine nucleotides and pyrimidines as a result of which HG type bps were formed (Makarova & Kulbachinskiy, 2012).

Hoogsteen pairings in DNA duplexes

DNA has a polymorphic nature, and its double helix can take different forms depending on the environment and sequence context. The earliest fiber X-ray diffraction studies highlighted this phenomenon. Further studies revealed that DNA is indeed made up of many different forms, as well as being capable of thermal fluctuations and large deformations in a sequence-dependent manner (Zhou, 2016). HG bps in duplex DNA can have a biological function, as shown by the biological studies showing it is important for cell survival in the presence of alkylating agents (Plosky et al., 2008) and oxidative stress (Petta et al., 2008). There was also considerable flexibility associated with the insipidly constrained sugar and phosphodiester backbones, as well as with the WC bps themselves (Travers, 2004). According to chemical probing and hydrogen exchange studies conducted between the 1970s and 1990s, WC bps opens at millisecond timescales with an abundance of less than 0.002 percent for AT and less than 0.00008 percent for GC bps in the open state (Englander et al., 1972). Several X-ray structures have now been constructed that arrest these open states of the bps when bound to proteins, which proves their functional significance. According to recent research, both A–T and G–C WC bps can transiently excursive toward HG bps in duplex DNA (Nikolova et al., 2012). The transient HG bps were categorized with the use of NMR R1ρ spectroscopic methods that make it conceivable to detect and anatomically illustrate evanescent states of macromolecules (Massi et al., 2004, Hansen et al., 2009). HG bps in the transient state had populations of 0.1 to 1%, almost three orders of magnitude greater than in the open state but were significantly less abundant at physiological pH than their A–T counterparts around the same pH range.

Compared to the forces in cells due to proteins, binding torsional stress, and supercoiling, or those generated by crystal packing forces, or those caused by pH and ionic changes, the differences between WC bp and HG bp are small (Zhou, 2016). In conclusion, the transient HG bps is found universally across all DNA sequence contexts, in a non-cooperative manner, and with small, but significant differences in population and lifetime across sequences (Alvey et al., 2014). It is apparent that every base pair in DNA exists in a rapid superposition of WC and HG bases, and external parameters operate on DNA to determine whether they are resolved by one or the other. WC versus HG has been observed for a long time, and the observation is controversial – minor deviations in conditions can favour one form over the other. As a result, the difference in the abundance of transient G–C^+ and A–T bps is striking since A/T replication is more efficient than G/C replication (Zhou, 2016). Adenine's higher N1 methylation frequency than guanine could be explained by the fact that the HG bps transiently expose the WC faces of purines. As a result of the observation of transient HG bps in duplex DNA, whose energetics are comparable to those of WC, it appears that HG bps might exist in much greater abundance in vivo, particularly in regions with high A–T content (Zhou, 2016). Given the current difficulties in separating WC from HG based on X-ray diffraction data, it is possible that more HG bps have gone undetected in X-ray structures currently deposited in Protein Data Bank, particularly for A–T bps.

Hoogsteen pairings in damaged DNA

A few external and endogenous factors could damage DNA by the 1960s, which in turn could contribute to diseases such as cancer (Alexander, 1954). As enzymes that recognize and repair damaged DNA became available in the 1970s and 1980s, there was a growing interest in characterizing damaged DNA (Friedberg, 2008). It was shown that HG base pairing could provide an important mechanism for hydrogen bonding and stacking, if the WC face of the purine bases is damaged, preventing favourable WC base pairing. According to Patel and colleagues, solution NMR studies showed for the first time that guanine adducts at the WC edge or C8 position favour syn base orientation which demarcated the first evidence for HG-type bps in damaged DNA in the late 1980s (Patel et al., 1986, Norman et al., 1989). NMR studies showed HG type pairing in

a number of purine lesions, including WC face alkylation adducts (such as 1,N2 propanoguanine and 1,N2 ethenoguanine and 1,N2 ethenoguanine), the bulky guanine C8 mutagenic adduct aminofluorine-C8 guanine and the common mutagenic lesion N1-methyladenine . In the 1990s and 2000s, HG base-pairing (rather than extra helical states) was directly observed in a wide range of naked DNA lesions, establishing HG base-pairing as a more energetically similar alternative to WC base-pairing (Zhou, 2016). There is much conjecture and experimental proof that HG-type pairs play a crucial role in DNA damage and mismatch repair. In pure-purine mismatches, flipping one purine base to a syn conformation is also often perceived, since the syn-anti conformation provides a smaller helical radius that can be accommodated within B-DNA more readily than the anti-anti form (Zhou, 2016).

There is an X-ray organizational indication that the DNA mismatch repair enzyme MutS explicitly recognizes HG type purine-purine and purine-pyrimidine mismatches despite not being the foremost conformation in unbound DNA, by creating definite hydrophobic and hydrogen bonding minor groove contact with the syn adenine/guanine base in A–C, A–A, and G–G mismatches (Natrajan et al., 2003). In certain mismatched bps, the enzyme may be able to discriminate against undamaged anti-anti WC bps by recognizing the increased population of syn-anti configuration rather than anti-anti configuration. Thus, HG bps not only provide a mechanism for maintaining the overall structural integrity of damaged or incorrectly replicated DNA but also contribute to DNA repair (Zhou, 2016). The addition of an ethylene bridge between C3’ and C5’ in bicyclo-DNA, which fixes the gamma backbone torsion angle to a noncanonical orientation, as well as HG bps has also been observed in DNA containing non-natural sugar-phosphate backbone modifications (Zhou, 2016). By substituting sugar O4’ with a methylene group in a single residue (Bolli et al., 1996), it fixes the gamma backbone torsion angle (Isaksson et al., 2001). At extremely low temperatures, dinucleotide d(TA) analogs contain a diisopropylsilyl-modified backbone (Bailor et al., 2010).

Hoogsteen pairings in RNA

It is not just Watson-Crick DNA that has a dominant structure, but RNA as well. According to Li et al, more than 50% of animal cell RNA is double-stranded, and double-stranded RNAs are widespread in functional RNAs including mRNAs, long non-coding RNAs, rRNAs, tRNAs, and transposable RNAs (Li et al., 2012). The basic structural building block of secondary, tertiary, and higher-order RNA structures is double-stranded RNA, which is often involved in RNA architectures, intermolecular interactions between RNA and RNA, including kissing dimers and mini-helixes formed by codons and anticodons (Zhou, 2016). The regulation of gene expression, translation, and RNA interference pathways are also influenced by double stranded RNAs (Cruz & Westhof, 2009). It has long been recognized that the canonical double helices formed by RNA (A-form) and DNA (B-form) differ in terms of their structure. There are several differences between the A-form helix and the B-form helix, including a higher rise and twist per bps, greater rolling, and displacement of bps away from the helical axis, wider helical diameter, and less and deeper major grooves. There has long been recognized that the canonical double helices formed by RNA (A-form) and DNA (B-form) differ in their structure (Zhou, 2016). As opposed to the B-form helix, the A-form helix produces a stiffer helix with lower rises and twists per bps step, a wider diameter, wider rolling, and displacement of bps away from the helical axis, a narrower major groove, and a larger helical diameter (Zhou, 2016). Due to steric contacts between the 2'- hydroxyl (OH) and 3'-O groups in RNA, the 2'-OH groups disfavour the C2'-endo sugar pucker that is preferred in B-DNA (Neidle, 2021). This is because of the electronegativity of the 2'-OH group and the physical properties of the C3'-endo sugar pucker (Brameld & Goddard, 1999). A-form helix is thereby compressed and rigidified, its helical diameter is widened as compared to B-form helix, and bps are displaced away from the helical axis due to proximity between O5' and O3' linking adjacent nucleotides. According to Felsenfeld et al., 1957, the observation of A–U and G–C WC bps in RNA created a false sense of comfort because recent studies have shown that, in contrast to B-form DNA, HG bases are unlikely to form in A-form RNA, as evidenced by the lack of HG bases in more than 1000 high-resolution crystal structures examined in the PDB for A-form RNA duplexes (Zhou, 2016). The A-form RNA double helix does not frequently contain HG bps, but they can occur in other structural contexts in RNA (Nagaswamy et al., 2002), at tertiary contacts in ribosomal RNA and transfer RNA, and in RNA triplexes; reverse rA–rU HG bps occur in these duplexes or triplexes (Nagaswamy et al., 2002). Different structural contexts can cause HG bps to have different structures, for example, HG or reverse HG bps in triplexes have anti rather than syn conformations; reverse A–U HG bps from one different H-bond (A-N6H---O2-U) compared to A–U HG bps (A-N6H---O4-U) and the C1’–C1’ distance (9.5 Å) is less constricted than that in HG bps (~ 8.5 Å) that arises due to the trans orientation of the nucleobases (Zhou, 2016).

Conclusion

Over the last few decades, DNA has been portrayed primarily as a right-handed B form double helix made of WC bps. The WC bps not only establishes a template mechanism for replication, transcription, and translation but also provides a basic structural building block for DNA and its interactions with proteins. With HG bps, the structure and chemical properties of the double helix can be significantly altered, thereby increasing its functional complexity. Bypassing replication damage, accommodating and repairing DNA damage, and recognizing DNA proteins are all important functions of HG bps. However,

it is unclear whether HG bps are widespread in nucleic acid duplexes and how they affect DNA and RNA. Many of these suspected HG bps are in AT-rich sequences or nucleosomes that feature sharp DNA bends. A major goal for the future is to examine HG bps in chromatin. Methods based on solid-state NMR and chemical probing can be used to investigate the occurrence of HG bps in potentially relevant in vivo environments. Further research is required to understand the forces that stabilize HG bps in a variety of contexts, including tight compaction and torsional stress in chromatin which generates an environment where HG bps are even more predominant. How much of these forces are related to the destabilization of WC BP compared with the stabilization of HG BP? Are stacking interactions a major contributor to these forces or are there other electrostatic effects at play?

In the past, studies have primarily focused on HG bps in duplex B-DNA. However recent studies investigated whether HG bps form in canonical A-form RNA duplexes as well. NMR RD performed over a broad range of sequence and structural contexts and under various environmental conditions revealed no evidence for transient HG bps in A-RNA duplexes. A-RNA duplexes are significantly destabilized because of the inability to form HG bps, which is the result of posttranslational modifications like m1A and m1G that prevent WC paring from occurring. A form of RNA has a higher affinity for melting than HG conformation, as it suppresses HG pairing so strongly that melting is preferred. However, DNA retains the ability to absorb damage including m1A and m1G modifications because HG pairing can be formed in HG bps. This results in a chemical switch in the form of m1rA and m1rG that can effectively control the structure, and thus the role, of the epitranscriptome. As a result, there is a basis for opposing functions at the genome and transcriptome levels because of the markedly different stabilities of the HG base. A lesion such as m1dA and m1dG that blocks canonical WC base pairing would greatly destabilize the double helix and potentially cause genomic instability if DNA had no capacity to form HG bps. There has been a recent finding that sugar pucker and phosphodiester backbones in A-RNA cause the HG bps to be disfavoured. It is also necessary to explore other factors including water interactions that may influence the relative stability of purine-purine HG bps in DNA and RNA and to confirm whether purine-purine HG bps are also more destabilized in A-RNA than in B-DNA (Johnson et al., 2005). In conclusion, in its reliance upon HG bps for DNA synthesis, Polι diverges from all other known DNA polymerases. It is proposed that this rare repetition mode grants Polι with the capability to integrate nucleotides contradictory to highly altering DNA lesions that invade the DNA minor groove or that disturb the W-C edge of the template purine (Zhou et al., 2005). In the review, one of the most noteworthy conclusions is that HG bps result in significant amounts of DNA bending (∼14◦) which is inversely proportional to the distance across the bp between C1'–C1'. Through indirect DNA sequence recognition, HG-mediated DNA bending may elucidate novel mechanisms.

References

Abrescia, N. G., González, C., Gouyette, C., & Subirana, J. A. (2004). X-ray and NMR studies of the DNA oligomer d(ATATAT): Hoogsteen base pairing in duplex DNA. Biochemistry, 43(14), 4092–4100. https://doi.org/10.1021/bi0355140.

Abrescia, N. G., Thompson, A., Huynh-Dinh, T., & Subirana, J. A. (2002). Crystal structure of an antiparallel DNA fragment with Hoogsteen base pairing. Proceedings of the National Academy of Sciences of the United States of America, 99(5), 2806–2811. https://doi.org/10.1073/pnas.052675499.

Acosta-Reyes, F. J., Alechaga, E., Subirana, J. A., & Campos, J. L. (2015). Structure of the DNA duplex d(ATTAAT)2 with Hoogsteen hydrogen bonds. PloS one, 10(3), e0120241. https://doi.org/10.1371/journal.pone.0120241.

Aishima, J., Gitti, R. K., Noah, J. E., Gan, H. H., Schlick, T., & Wolberger, C. (2002). A Hoogsteen base pair embedded in undistorted B-DNA. Nucleic acids research, 30(23), 5244–5252. https://doi.org/10.1093/nar/gkf661.

Alexander, P. (1954). Advances in Cancer Research.Volume 2 (eds P. Greenstein Jesse & Haddow Alexander) 1-72. Academic Press

Nikolova, E. N., Kim, E., Wise, A. A., O'Brien, P. J., Andricioaei, I., & Al-Hashimi, H. M. (2011). Transient Hoogsteen base pairs in canonical duplex DNA. Nature, 470(7335), 498–502. https://doi.org/10.1038/nature09775.

Atul Rangadurai, Honglue Shi, Yu Xu, Bei Liu, Hala Abou Assi, John D. Boom, Huiqing Zhou, Isaac J. Kimsey, Hashim M. Al-Hashimi, (2022). Measuring thermodynamic preferences to form non-native conformations in nucleic acids using ultraviolet melting, Proceedings of the National Academy of Sciences, 119, 24. doi/10.1073/pnas.2112496119

Bailor, M. H., Sun, X., & Al-Hashimi, H. M. (2010). Topology links RNA secondary structure with global conformation, dynamics, and adaptation. Science (New York, N.Y.), 327(5962), 202–206. https://doi.org/10.1126/science.1181085.

Blommers, M.J.J., Van De Ven, F.J.M., Van Der Marel, G.A., Van Boom, J.H. and Hilbers, C.W. (1991), The three-dimensional structure of a DNA hairpin in solution. European Journal of Biochemistry, 201: 33-51. https://doi.org/10.1111/j.1432-1033.1991.tb16253.x.

Bolli, M., Litten, J. C., Schütz, R., & Leumann, C. J. (1996). Bicyclo-DNA: a Hoogsteen-selective pairing system. Chemistry & biology, 3(3), 197–206. https://doi.org/10.1016/s1074-5521(96)90263-x.

Brameld, K. A. & Goddard, W. A. (1999) Ab initio quantum mechanical study of the structures and energies for the pseudorotation of 5′-dehydroxy analogues of 2′-deoxyribose and ribose sugars. J. Am. Chem. Soc. 121, 985-993. https://doi.org/10.1021/ja982995f.

Chen, Y., Dey, R., & Chen, L. (2010). Crystal structure of the p53 core domain bound to a full consensus site as a self-assembled tetramer. Structure (London, England : 1993), 18(2), 246–256. https://doi.org/10.1016/j.str.2009.11.011.

Cruz, J. A., & Westhof, E. (2009). The dynamic landscapes of RNA architecture. Cell, 136(4), 604–609. https://doi.org/10.1016/j.cell.2009.02.003.

DAVIES, D. R., & BALDWIN, R. L. (1963). X-ray studies on two synthetic DNA copolymers. Journal of molecular biology, 6, 251–255. https://doi.org/10.1016/s0022-2836(63)80086-8.
Drew, H. R., & Dickerson, R. E. (1982). A new model for DNA containing A.T and I.C base pairs. The EMBO journal, 1(6), 663–667. https://doi.org/10.1002/j.1460-2075.1982.tb01227.x.
Englander, S. W., Downer, N. W., & Teitelbaum, H. (1972). Hydrogen exchange. Annual review of biochemistry, 41, 903–924. https://doi.org/10.1146/annurev.bi.41.070172.004351.
FELSENFELD, G., & RICH, A. (1957). Studies on the formation of two- and three-stranded polyribonucleotides. Biochimica et biophysica acta, 26(3), 457–468. https://doi.org/10.1016/0006-3002(57)90091-4.
Friedberg E. C. (2008). A brief history of the DNA repair field. Cell research, 18(1), 3–7. https://doi.org/10.1038/cr.2007.113.
Hansen, A. L., Nikolova, E. N., Casiano-Negroni, A., & Al-Hashimi, H. M. (2009). Extending the range of microsecond-to-millisecond chemical exchange detected in labeled and unlabeled nucleic acids by selective carbon R(1rho) NMR spectroscopy. Journal of the American Chemical Society, 131(11), 3818–3819. https://doi.org/10.1021/ja8091399.
Hoogsteen K. (1959) The structure of crystals containing a hydrogen-bonded complex of 1-methylthymine and 9-methyladenine. Acta crystallographica ;12(10):822–823. https://doi.org/10.1107/S0365110X59002389.
Hoopes, B. C., LeBlanc, J. F., & Hawley, D. K. (1998). Contributions of the TATA box sequence to rate-limiting steps in transcription initiation by RNA polymerase II. Journal of molecular biology, 277(5), 1015–1031. https://doi.org/10.1006/jmbi.1998.1651.
Ikehara, Morio & Hattori, Masao & Fukui, Toshikazu. (2005). Synthesis and Properties of Poly(2-Methyladenylic Acid). Formation of a Poly(A) . Poly(U) Complex with Hoogsteen-Type Hydrogen Bonding. European Journal of Biochemistry. 31. 329 - 334. https://doi.org/10.1111/j.1432-1033.1972.tb02537.x.
Ben Imeddourene, A., Zargarian, L., Buckle, M., Hartmann, B., & Mauffret, O. (2020). Slow motions in A·T rich DNA sequence. Scientific reports, 10(1), 19005. https://doi.org/10.1038/s41598-020-75645-x.
Isaksson, J., Zamaratski, E., Maltseva, T. V., Agback, P., Kumar, A., & Chattopadhyaya, J. (2001). The first example of a Hoogsteen base-paired DNA duplex in dynamic equilibrium with a Watson-Crick base-paired duplex--a structural (NMR), kinetic and thermodynamic study. Journal of biomolecular structure & dynamics, 18(6), 783–806. https://doi.org/10.1080/07391102.2001.10506707.
Ishikawa, F., Frazier, J., Howard, F. B., & Miles, H. T. (1972). Polyadenylate polyuridylate helices with non-Watson-Crick hydrogen bonding. Journal of molecular biology, 70(3), 475–490. https://doi.org/10.1016/0022-2836(72)90554-2.
Jocelyne Vreede, Alberto Pérez de Alba Ortíz, Peter G Bolhuis, David W H Swenson. (2019). Atomistic insight into the kinetic pathways for Watson–Crick to Hoogsteen transitions in DNA, Nucleic Acids Research, 47, 21, (11069-11076). https://doi.org/10.1093/nar/gkz837.

Johnson, R. E., Prakash, L., & Prakash, S. (2005). Biochemical evidence for the requirement of Hoogsteen base pairing for replication by human DNA polymerase iota. Proceedings of the National Academy of Sciences of the United States of America, 102(30), 10466–10471. https://doi.org/10.1073/pnas.0503859102.
Kitayner, M., Rozenberg, H., Rohs, R., Suad, O., Rabinovich, D., Honig, B., & Shakked, Z. (2010). Diversity in DNA recognition by p53 revealed by crystal structures with Hoogsteen base pairs. Nature structural & molecular biology, 17(4), 423–429. https://doi.org/10.1038/nsmb.1800.
LEHMAN, I. R., BESSMAN, M. J., SIMMS, E. S., & KORNBERG, A. (1958). Enzymatic synthesis of deoxyribonucleic acid. I. Preparation of substrates and partial purification of an enzyme from Escherichia coli. The Journal of biological chemistry, 233(1), 163–170.
Li, F., Zheng, Q., Ryvkin, P., Dragomir, I., Desai, Y., Aiyer, S., Valladares, O., Yang, J., Bambina, S., Sabin, L. R., Murray, J. I., Lamitina, T., Raj, A., Cherry, S., Wang, L. S., & Gregory, B. D. (2012). Global analysis of RNA secondary structure in two metazoans. Cell reports, 1(1), 69–82. doi: 10.1016/j.celrep.2011.10.002]
Ling, H., Boudsocq, F., Plosky, B. S., Woodgate, R., & Yang, W. (2003). Replication of a cis-syn thymine dimer at atomic resolution. Nature, 424(6952), 1083–1087. https://doi.org/10.1038/nature01919.
Liu, K., Miles, H. T., Frazier, J., & Sasisekharan, V. (1993). A novel DNA duplex. A parallel-stranded DNA helix with Hoogsteen base pairing. Biochemistry, 32(44), 11802–11809. https://doi.org/10.1021/bi00095a008.
MacKerell, A. D., Jr, Banavali, N., & Foloppe, N. (2000). Development and current status of the CHARMM force field for nucleic acids. Biopolymers, 56(4), 257–265. https://doi.org/10.1002/1097-0282(2000)56:4<257::AID-BIP10029>3.0.CO;2-W.
Makarova, A. V., & Kulbachinskiy, A. V. (2012). Structure of human DNA polymerase iota and the mechanism of DNA synthesis. Biochemistry. Biokhimiia, 77(6), 547–561. https://doi.org/10.1134/S0006297912060016.
Malecka, K. A., Ho, W. C., & Marmorstein, R. (2009). Crystal structure of a p53 core tetramer bound to DNA. Oncogene, 28(3), 325–333. https://doi.org/10.1038/onc.2008.400.
Massi, F., Johnson, E., Wang, C., Rance, M., & Palmer, A. G., 3rd (2004). NMR R1 rho rotating-frame relaxation with weak radio frequency fields. Journal of the American Chemical Society, 126(7), 2247–2256. https://doi.org/10.1021/ja038721w.
Meyer, T., Gustafsson, J. A., & Carlstedt-Duke, J. (1997). Glucocorticoid-dependent transcriptional repression of the osteocalcin gene by competitive binding at the TATA box. DNA and cell biology, 16(8), 919–927. https://doi.org/10.1089/dna.1997.16.919.

Nagaswamy, U., Larios-Sanz, M., Hury, J., Collins, S., Zhang, Z., Zhao, Q., & Fox, G. E. (2002). NCIR: a database of non-canonical interactions in known RNA structures. Nucleic acids research, 30(1), 395–397. https://doi.org/10.1093/nar/30.1.395.
Nair, D. T., Johnson, R. E., Prakash, S., Prakash, L., & Aggarwal, A. K. (2004). Replication by human DNA polymerase-iota occurs by Hoogsteen base-pairing. Nature, 430(6997), 377–380. https://doi.org/10.1038/nature02692.
Natrajan, G., Lamers, M. H., Enzlin, J. H., Winterwerp, H. H., Perrakis, A., & Sixma, T. K. (2003). Structures of Escherichia coli DNA mismatch repair enzyme MutS in complex with different mismatches: a common recognition mode for diverse substrates. Nucleic acids research, 31(16), 4814–4821. https://doi.org/10.1093/nar/gkg677.
Neidle, S., & Sanderson, M. (2021). Principles of nucleic acid structure. Academic Press.
Nikolova, E. N., Kim, E., Wise, A. A., O'Brien, P. J., Andricioaei, I., & Al-Hashimi, H. M. (2011). Transient Hoogsteen base pairs in canonical duplex DNA. Nature, 470(7335), 498–502. https://doi.org/10.1038/nature09775.
Nikolova, E. N., Zhou, H., Gottardo, F. L., Alvey, H. S., Kimsey, I. J., & Al-Hashimi, H. M. (2013). A historical account of Hoogsteen base-pairs in duplex DNA. Biopolymers, 99(12), 955–968. https://doi.org/10.1002/bip.22334.
Nikolova, E. N., Gottardo, F. L., & Al-Hashimi, H. M. (2012). Probing transient Hoogsteen hydrogen bonds in canonical duplex DNA using NMR relaxation dispersion and single-atom substitution. Journal of the American Chemical Society, 134(8), 3667–3670. https://doi.org/10.1021/ja2117816.
Norman, D., Abuaf, P., Hingerty, B. E., Live, D., Grunberger, D., Broyde, S., & Patel, D. J. (1989). NMR and computational characterization of the N-(deoxyguanosin-8-yl)aminofluorene adduct [(AF)G] opposite adenosine in DNA: (AF)G[syn].A[anti] pair formation and its pH dependence. Biochemistry, 28(18), 7462–7476. https://doi.org/10.1021/bi00444a046.
Ol'ha O. Brovarets', Kostiantyn S. Tsiupa, Dmytro M. Hovorun, (2021).Where Quantum Biochemistry Meets Structural Bioinformatics: Excited Conformationally-Tautomeric States of the Classical A·T DNA Base Pair, DNA - Damages and Repair Mechanisms. https://doi.org/10.5772/intechopen.87549
Patel, D. J., Shapiro, L., Kozlowski, S. A., Gaffney, B. L., & Jones, R. A. (1986). Covalent carcinogenic O6-methylguanosine lesions in DNA. Structural studies of the O6 meG X A and O6meG X G interactions in dodecanucleotide duplexes. Journal of molecular biology, 188(4), 677–692. https://doi.org/10.1016/s0022-2836(86)80014-6.
Patikoglou, G. A., Kim, J. L., Sun, L., Yang, S. H., Kodadek, T., & Burley, S. K. (1999). TATA element recognition by the TATA box-binding protein has been conserved throughout evolution. Genes & development, 13(24), 3217–3230. https://doi.org/10.1101/gad.13.24.3217.
Pérez, A., Marchán, I., Svozil, D., Sponer, J., Cheatham, T. E., 3rd, Laughton, C. A., & Orozco, M. (2007). Refinement of the AMBER force field for nucleic acids: improving the description of alpha/gamma conformers. Biophysical journal, 92(11), 3817–3829. https://doi.org/10.1529/biophysj.106.097782.
Pérez de Alba Ortíz, A., Vreede, J., & Ensing, B. (2022). Sequence dependence of transient Hoogsteen base pairing in DNA. PLoS computational biology, 18(5), e1010113. https://doi.org/10.1371/journal.pcbi.1010113.
Petta, T. B., Nakajima, S., Zlatanou, A., Despras, E., Couve-Privat, S., Ishchenko, A., Sarasin, A., Yasui, A., & Kannouche, P. (2008). Human DNA polymerase iota protects cells against oxidative stress. The EMBO journal, 27(21), 2883–2895. doi: 10.1038/emboj.2008.210.
Plosky, B. S., Frank, E. G., Berry, D. A., Vennall, G. P., McDonald, J. P., & Woodgate, R. (2008). Eukaryotic Y-family polymerases bypass a 3-methyl-2'-deoxyadenosine analog in vitro and methyl methanesulfonate-induced DNA damage in vivo. Nucleic acids research, 36(7), 2152–2162. https://doi.org/10.1093/nar/gkn058.
Ray, D., & Andricioaei, I. (2020). Free Energy Landscape and Conformational Kinetics of Hoogsteen Base Pairing in DNA vs. RNA. Biophysical journal, 119(8), 1568–1579. https://doi.org/10.1016/j.bpj.2020.08.031.
Rice, P. A., Yang, S., Mizuuchi, K., & Nash, H. A. (1996). Crystal structure of an IHF-DNA complex: a protein-induced DNA U-turn. Cell, 87(7), 1295–1306. https://doi.org/10.1016/s0092-8674(00)81824-3.
Robert J. Fick, Amy Y. Liu, Felix Nussbaumer, Christoph Kreutz, Atul Rangadurai, Yu Xu, Roger D. Sommer, Honglue Shi, Steve Scheiner, Allison L. Stelling.(2021).Probing the Hydrogen-Bonding Environment of Individual Bases in DNA Duplexes with Isotope-Edited Infrared Spectroscopy, The Journal of Physical Chemistry B, 125, 28, (7613-7627).https://doi.org/10.1021/acs.jpcb.1c01351
Ronning, D. R., Guynet, C., Ton-Hoang, B., Perez, Z. N., Ghirlando, R., Chandler, M., & Dyda, F. (2005). Active site sharing and subterminal hairpin recognition in a new class of DNA transposases. Molecular cell, 20(1), 143–154. doi: 10.1016/j.molcel.2005.07.02
Segers-Nolten, G. M., Sijtsema, N. M., & Otto, C. (1997). Evidence for Hoogsteen GC base pairs in the proton-induced transition from right-handed to left-handed poly(dG-dC).poly(dG-dC). Biochemistry, 36(43), 13241–13247. https://doi.org/10.1021/bi971326w.
Song, K., Campbell, A. J., Bergonzo, C., de Los Santos, C., Grollman, A. P., & Simmerling, C. (2009). An Improved Reaction Coordinate for Nucleic Acid Base Flipping Studies. Journal of chemical theory and computation, 5(11), 3105–3113. https://doi.org/10.1021/ct9001575.
Tissier, A., McDonald, J. P., Frank, E. G., & Woodgate, R. (2000). poliota, a remarkably error-prone human DNA polymerase. Genes & development, 14(13), 1642–1650.
Travers A. A. (2004). The structural basis of DNA flexibility. Philosophical transactions. Series A, Mathematical, physical, and engineering sciences, 362(1820), 1423–1438. https://doi.org/10.1098/rsta.2004.1390.

Wang J. (2005). DNA polymerases: Hoogsteen base-pairing in DNA replication?. Nature, 437(7057), E6–E7. https://doi.org/10.1038/nature04199.
WATSON, J. D., & CRICK, F. H. (1953). Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. Nature, 171(4356), 737–738. https://doi.org/10.1038/171737a0.
Yang, C., Kim, E., & Pak, Y. (2015). Free energy landscape and transition pathways from Watson-Crick to Hoogsteen base pairing in free duplex DNA. Nucleic acids research, 43(16), 7769–7778. https://doi.org/10.1093/nar/gkv796.
Zhou, H., Hintze, B. J., Kimsey, I. J., Sathyamoorthy, B., Yang, S., Richardson, J. S., & Al-Hashimi, H. M. (2015). New insights into Hoogsteen base pairs in DNA duplexes from a structure-based survey. Nucleic acids research, 43(7), 3420–3433. https://doi.org/10.1093/nar/gkv241.
Zhou, Huiqing (2016). Occurrence and Function of Hoogsteen Base Pairs in Nucleic Acids. Dissertation, Duke University.