

Published in final edited form as:

Biopolymers. 2013 December; 99(12): . doi:10.1002/bip.22334.

# A Historical Account of Hoogsteen Base-pairs in Duplex DNA

Evgenia N. Nikolova<sup>1,2</sup>, Huiqing Zhou<sup>1</sup>, Federico L. Gottardo<sup>1</sup>, Heidi S. Alvey<sup>1</sup>, Isaac J. Kimsey<sup>1</sup>, and Hashim M. Al-Hashimi<sup>1,\*</sup>

<sup>1</sup>Department of Chemistry & Biophysics, The University of Michigan, 930 North University Avenue, Ann Arbor, MI 48109-1055, USA, Tel: 734-615-3361, Fax: 734-647-4865

<sup>2</sup>Integrative Structural & Computational Biology Department, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, CA 92037, USA

#### **Abstract**

In 1957, a unique pattern of hydrogen bonding between N3 and O4 on uracil and N7 and N6 on adenine was proposed to explain how poly(rU) strands can associate with poly(rA)-poly(rU) duplexes to form triplexes. Two years later, Karst Hoogsteen visualized such a non-canonical A-T base-pair through X-ray analysis of co-crystals containing 9-methyladenine and 1-methylthymine. Subsequent X-ray analyses of guanine and cytosine derivatives yielded the expected Watson-Crick base-pairing but those of adenine and thymine (or uridine) did not yield Watson-Crick base-pairs, instead favoring 'Hoogsteen' base-pairing. More than two decades ensued without experimental 'proof' for A-T Watson-Crick base-pairs, while Hoogsteen base-pairs continued to surface in ATrich sequences, closing base-pairs of apical loops, in structures of DNA bound to antibiotics and proteins, damaged and chemically modified DNA, and in polymerases that replicate DNA via Hoogsteen pairing. Recently, NMR studies have shown that base-pairs in duplex DNA exists as a dynamic equilibrium between Watson-Crick and Hoogsteen forms. There is now little doubt that Hoogsteen base-pairs exist in significant abundance in genomic DNA where they can expand the structural and functional versatility of duplex DNA beyond that which can be achieved based only on Watson-Crick base-pairing. Here, we provide a historical account of the discovery and characterization of Hoogsteen base-pairs hoping that this will inform future studies exploring the occurrence and functional importance of these alternative base-pairs.

#### Introduction

In 1953, sixty years ago, Watson and Crick proposed their iconic double helix structure for deoxyribonucleic acid (DNA) based on very little experimental data. Although the structure is most known for its double helical appearance, its most important feature was, and remains to this date, the specific pairing of purine with pyrimidine nucleobases – guanine with cytosine and adenine with thymine – through complimentary hydrogen bonds (Figure 1). This endowed the structure with the ability to self-duplicate, making DNA, and not proteins as was widely believed at the time, the likely carrier of genetic information. Despite the absence of any experimental data in support of the specific pairing proposed by Watson and Crick, and despite the fact that there are alternative modes for pairing purines with pyrimidines, the pairing proposed by Watson and Crick utilized bases in their most probable tautomeric forms and most importantly, resulted in similar overall shapes for all four basepair combinations, so that any sequence could be accommodated within the same double helix framework.

<sup>\*</sup>hashimi@umich.edu.

Although the discovery of the double helix set in motion one of the greatest scientific revolutions, the structure itself was met with a good deal of skepticism. The available X-ray fiber diffraction data obtained on non-crystalline DNA fibers, particularly B-form DNA, did not provide adequate resolution to determine atomic positions. This is because molecules in the fiber are generally not rotationally oriented relative to one another in a regular manner. Indeed, this was the main reason Rosalind Franklin pursued the more complicated diffraction pattern presented by the dry "A-form" version of DNA, here the molecules are not in random rotational orientations, allowing for a more objective 3D crystallographic analysis and where one could, in Franklin's own words, 'let the data speak for itself'.

While the story of the double helix is well known to scientists and non-scientists alike, it is not commonly known that definitive proof for the DNA double helix structure did not come until 1980 – more than a quarter century after Watson and Crick initially proposed their model – when Drew, Dickerson and co-workers solved the single crystal structure of a DNA dodecamer using heavy atom X-ray crystallography.<sup>5,6</sup> In the ensuing period, experimental evidence began to accumulate for an alternative base-pair, referred to now as the 'Hoogsteen' base-pair (Figure 1),<sup>7,8</sup> which, together with other alternative structures of DNA such as left-handed Z-DNA,<sup>9</sup> raised doubts about the B-form structure proposed by Watson and Crick. Today, there is little doubt that Hoogsteen (HG) base-pairs do indeed represent an alternative pairing scheme that can expand the structural and functional versatility of duplex DNA beyond that which can be achieved based only on Watson-Crick base-pairing. The purpose of this review is to provide a historical account of the discovery and characterization of HG base-pairs hoping that this will inform future studies exploring the occurrence and functional importance of these alternative base-pairs.

# Purine-pyrimidine co-crystals

Soon after Watson and Crick proposed their double helix structure, experimentalists rushed to gather data to test its various aspects. Improvements in X-ray cameras and analytical methods for refining models to fit X-ray fiber diffraction data in the following years provided additional evidence in support of the general features of the DNA double helix model. However, fine details of the structure, including the specific base-pairs proposed by Watson and Crick, could not be assessed. Throughout the late 1950s and into the 1970s, much effort was directed toward solving X-ray structures of isolated purine-pyrimidine dimers. The idea was that the monomers might associate to form intermolecular complexes that reflect the pairing that occurs in the double helix. The high-resolution diffraction data afforded by single crystals allowed for an objective characterization of hydrogen bonding interactions between the bases.

The first such study was reported in 1959, when Karst Hoogsteen – an associate of Robert Corey at Caltech –used single crystal X-ray analysis to determine the structures of co-crystals containing 9-methyladenine and 1-methylthymine, where methyl groups were used to block hydrogen bonding to nitrogen atoms otherwise bonded to sugar carbons in DNA. Rather than observing a Watson-Crick (WC) base-pair, Hoogsteen observed a markedly different pairing scheme (Figure 1), in which the adenine base was flipped upside down. In DNA, such a flip is accomplished by a 180-degree rotation of the adenine base around the glycosidic bond (N9—C1), changing the base from an *anti* to a *syn* conformation. As in WC base-pairs, the thymine base formed two hydrogen bonds with the adenine base, one of which (thymine O4 and adenine N6) is identical to that proposed by Watson and Crick. However, the second hydrogen bond is not between thymine N3 and adenine N1, but rather, between thymine N3 and N7 of the flipped adenine base (Figure 1). This very same hydrogen bonding scheme was proposed two years earlier by Rich and his colleagues to explain how poly(rU) strands might associate with poly(rA)-poly(rU) duplexes to form

triplexes. <sup>10</sup> Hoogsteen recognized that relative to the scheme proposed by Watson and Crick, this hydrogen bonding scheme required translation of the complementary bases into closer proximity, which requires constriction of the DNA helix diameter by  $\sim 2.5 \text{ Å}.^{7.8}$ 

In 1963, one year after Watson and Crick were awarded the Noble Prize in Physiology or Medicine, single crystal structures were reported for G-C base-pairs using crystals containing 9-ethylguanine and 1-methylcytosine or 1-methyl-5-bromocytosine. WC base-pairing was observed in the two cases, even though the two intermolecular complexes crystallized in different space groups and experienced different packing arrangements. As predicted earlier by Pauling and his colleagues that G-C base-pairs were stabilized by three and not two hydrogen bonds as proposed by Watson and Crick. Subsequent X-ray diffraction analysis of co-crystals involving guanine and cytosine derivatives consistently formed the expected WC base-pairs. However, in sharp contrast, attempts to generate co-crystals of adenines and thymine (or uridine) derivatives failed to yield WC base-pairs and, in most cases, favored 'Hoogsteen' (HG) base-pairs. 13–15

In 1968, Guschlbauer and colleagues proposed the formation of G-C<sup>+</sup> HG base-pairs in poly(dG)-poly(dC) at pH 3–4 based on optical rotatory dispersion spectra suggesting guanine adopted a *syn* conformation. <sup>16</sup> As in A-T base-pairs, the transition from WC to HG entails flipping of the guanine base from an *anti* to a *syn* conformation and the preservation of hydrogen bonding between cytosine N4 and guanine O6. However, unlike A-T base-pairs, formation of a second hydrogen bond with N7 of the flipped guanine requires protonation of cytosine N3. The transition results in a net loss of one hydrogen bond and the build-up of a positive charge on the cytosine. Because of this, G-C<sup>+</sup> HG base-pairs were generally thought to be less energetically favored than A-T HG base-pairs and more stably formed at lower pH. <sup>17</sup> G-C<sup>+</sup> HG base-pair were subsequently used to explain how poly(rC) associates with poly(dC)-poly(dG) duplexes to form triplexes under acidic conditions <sup>18</sup> and NMR studies provided chemical shift evidence for protonated G-C<sup>+</sup> HG base-pairs at cytosine N3 in a poly(dC)-poly(dC) complex with dGMP at low pH. <sup>19</sup> The first crystallographic observation of G-C<sup>+</sup> HG base-pair came many years later for a DNA duplex bound to the bisintercalating antibiotic triostin A. <sup>20</sup>

HG base-pairs presented an alternative to WC base-pairs, especially for A-T rich sequences, and this contributed to skepticism about the details of the DNA double helix structure. Among the skeptics was Linus Pauling, who had proposed an incorrect triple helix model for the DNA structure the same year that Watson and Crick proposed their double helix model. Indeed, Pauling felt that awarding Watson and Crick a Noble Prize for their discovery was premature "because of existing uncertainty about the detailed structure of nucleic acid" (personal correspondence to the Nobel Committee for Chemistry and Physics). The only real experimental indication that both G-C and A-T/U could form WC base-pairs in double and triple helices came from subsequent solution NMR studies of tRNA and polynucleotide complexes in 1970s. 19,22–27 These studies, which were performed under physiological solution conditions in the absence of potentially perturbing crystal packing forces, showed distinct chemical shift signatures that were consistent with theoretical predictions for WC rather than HG hydrogen bonding in A-T/U base-pairs.

# **HG Base-pairs in Naked Duplexes**

In 1973, Rich and colleagues reported the single crystal X-ray structure of the AU and GC dinucleoside phosphates and the results were heralded as 'the double helix at atomic resolution'. <sup>28,29</sup> These structures verified key aspects of the double helix proposed by Watson and Crick. Both structures revealed a right-handed double helix with two strands running anti-parallel to each other. Importantly, both structures featured WC type base-

pairing. This was the first time a WC rather than HG base-pair was observed involving adenine. The structures seemed to put the controversy regarding HC versus WC A-T base-pairs to rest; since it was the WC form that was favored when the bases were constrained in a double helix. As noted by Alex Rich, James Watson phoned him after receiving a preprint of the AU manuscript and said "he had his first good night sleep in 20 years!" In retrospect, the observation of A-U WC base-pairs in A-form RNA created a false sense of comfort as recent studies have shown that, in contrast to DNA, HG base-pairs are not likely to form in A-form RNA, as suggested by the lack of HG base-pairs in A-form RNA duplexes in over one thousand high resolution crystal structures surveyed in the PDB.

Six years later, in 1979, following advances in phosphotriester methods for chemical synthesis of large quantities of homogeneous oligonucleotides, \$^{31,32}\$ Rich and colleagues reported the first single crystal X-ray structure of DNA for the d(CG)<sub>3</sub> sequence.9 Prior studies had shown that repeating polymers of inosine-cytosine<sup>33</sup> and guanine-cytosine<sup>34</sup> resulted in a "reverse" circular dichroism (CD) spectrum and this was interpreted as evidence for a left-handed helix. The structure of d(CG)<sub>3</sub> revealed a stunning left-handed double helix with an unusual zig-zag shape and was called "Z-DNA". To convert B-DNA into Z-DNA, both bases in the base-pair have to be flipped upside down. As in HG base-pairs, the guanine base is flipped into a *syn* conformation; however, in Z-DNA, the concomitant flipping of the cytosine base and sugar allows the two flipped bases to regroup into WC base-pairs, with the flipping of the cytosine sugar giving rise to the unusual zig-zag backbone. Thus, although the first single crystal structure of DNA provided evidence for G-C WC base-pairs, it fueled skepticism about the overall structure of the double helix.

In the early 1980s, models were put forward for a Z-DNA structure that are exclusively comprised of HG base-pairs,<sup>35</sup> particularly for A-T rich sequences that frequently exhibited unusual diffraction patterns when dried (referred to as D- or E- type X-ray diffraction patterns).<sup>35–37</sup> This form of DNA required helical structures with 7–7.5 base-pairs per turn, which cannot be stereochemically achieved by right-handed B-form DNA. Spectroscopic studies of poly(rA)-poly(rU) sequences that bear substituent at the adenine C2 position that sterically block WC base-pairing also suggested formation of duplexes with parallel or antiparallel chain polarity, in which strands are held together by HG or reverse HG base-pairing, respectively.<sup>38–40</sup>

During the same period, in 1980, Drew and Dickerson reported the first single crystal X-ray structure of a DNA polymer longer than one helical turn. They used heavy atom replacement approaches to solve the structure of a synthetic DNA dodecamer sequence d(CGCGAATTCGCG). The structure adopted a right-handed duplex containing WC G-C and A-T base-pairs – precisely as proposed by Watson and Crick. It is this structure that is considered to be crystallographic proof that DNA can indeed adopt the structure proposed by Watson and Crick.

The ability to prepare large quantities of highly pure DNA samples in a facile manner, in parallel with developments in \$^{13}\text{C}/^{15}\text{N}\$ isotopic enrichment and solution state NMR spectroscopy of nucleic acids resulted in the high-resolution X-ray and NMR structure determination of diverse DNA sequences in the 1980s and 1990s showing WC B-form DNA duplexes. However, spectroscopic evidence for HG base-pairs continued to mount in the 1990s and 2000s in the context of A-T rich sequences, \$^{41}\$ in poly(dG-dC)-poly(dG-dC) sequences at low pH as possible intermediates along the B-to-Z DNA transition, \$^{42}\$ as well as in non-canonical DNA regions as closing base-pairs of apical loops. \$^{43,44}\$ But it was not until 2002, when Subirana and colleagues reported the first crystal structure of an AT-repeat not capped by GC base-pairs, that the first single crystal X-ray structure of a naked DNA duplex containing exclusively HG base-pairs was resolved. \$^{45}\$ The structure of d(AT)3 revealed an

anti-parallel right-handed double helix made up exclusively of HG base-pairs with an overall structure similar to that of B-form DNA (Figure 2).  $^{45,46}$  Key differences included a change in the position of the helical axis relative to the base-pairs, reduction in helical radius and C1 -C1 distance by ~2.5–3.0 Å, altered hydrogen bonding donor/acceptor pattern in the major and minor grooves, and a narrower and less electronegative, minor groove which favors hydrophobic interactions and distinct helix stacking and hydration patterns relative to B-DNA. Together, these features provide a distinct physicochemical presentation of the genetic code for potential sequence-specific recognition by the cellular machinery. Similar HG structures were subsequently reported for related sequences d(ATATATCT) $^{47}$  and d(CGATATATATAT).

It is important to note that in all cases, solution state NMR studies of the above DNA sequences under the same conditions used to grow crystals argued against formation of a HG helix, and in favor of a prototypical WC B-form double helix (Zhou *et al.* unpublished data). <sup>46</sup> This suggests that crystal packing plays an important role in stabilizing the HG double helix.

## **DNA-antibiotic complexes**

In 1984, Rich and colleagues reported the single crystal X-ray structure of a DNA double helix with sequence d(CGTACG) bound to triostin A, <sup>49</sup> a cyclic octadepsipeptide anti-tumor antibiotic containing two quinoxaline rings that binds DNA and inhibits replication and transcription in vivo (Figure 3).<sup>50,51</sup> This was the first structure of a peptide antibiotic in complex with an oligonucleotide. The structure showed that the two quinoxaline rings bisintercalate in the minor groove of the DNA double helix and surround the WC G-C basepairs, disrupting stacking interactions to the central A-T base-pairs (Figure 3). Remarkably, although the two central A-T base-pairs are not covered by the two triostin A molecules, they form HG rather than the WC base-pairs. This marked the first crystallographic observation of the co-existence of WC and HG base-pairs within the same duplex. No direct contacts are observed between the antibiotic and the exposed Watson-Crick face of the A-T bases. Rather, the helical constriction at the HG base-pairs appears to stabilize the complex by allowing close packing of the oligonucleotide around the end of the triostin A. Thus, several favorable van der Waals contacts would be lost if the deoxyribose rings were further apart as in WC base-pairs. Similar structures were subsequently reported for DNA bound to the related echinomycin antibiotic, <sup>52</sup> and for triostin A bound to d(GCGTACGC), <sup>20</sup> which featured two central A-T HG base-pairs, and two terminal G-C+ HG base-pairs, marking the first crystallographic observation of protonated G-C+ HG base-pairs within a duplex (Figure 3).

Soon after, chemical footprinting studies performed in solution showed that sites that form HG base-pairs in X-ray structures of DNA-echinomycin complexes are hyperreactive to diethyl pyrocarbonate (DEPC),<sup>53</sup> which preferentially reacts with exposed N7 atoms of *syn* purines in non-canonical Z-DNA<sup>54</sup> and cruciform loops.<sup>55</sup> However, these results were challenged by footprinting studies employing DEPC and other reagents that target thymines, that showed little change in thymine chemical reactivity when replacing adenine with 7-deazaadenine, which has a diminished ability to form HG base-pairs.<sup>56–59</sup> Moreover, oligonucleotides containing 7-deazaadenine and 7-deazaguanine bound echinomycin with affinity comparable to that of their unmodified counterparts, suggesting that HG base-pairs are not essential for binding.<sup>59</sup> These studies argued that hyperreactivity does not arise from formation of HG base-pairs but rather from unwinding and extension of the DNA helix upon drug binding.

Subsequent NMR studies by Feigon, Patel, and their co-workers confirmed formation of HG base-pairs in DNA-antibiotic complexes,  $^{60,61}$  although their occurrence was shown to be highly dependent on sequence, temperature, and pH.  $^{62-64}$  A-T base-pairs generally form WC, and if they ever form HG, they do so transiently at physiological temperatures. Even the terminal HG A-T base-pairs were only favored in DNA-antibiotic complexes having purine 5 and pyrimidine 3 to CG (i.e. ACGT, GCGC) and only at low pH for G-C+ HG base-pairs.  $^{60,62,65,66}$ 

Despite many studies, to date, it remains unclear whether quinoxaline antibiotics stabilize HG DNA base-pairs *in vivo* and whether this is related in any way to their biological activity.

## **DNA-Protein complexes**

In the late 1990s, X-ray structures emerged showing that certain proteins bind and in some cases specifically recognize HG base-pairs embedded in B-form DNA (Figure 4). These studies raised the possibility that proteins exploit the unique structural and chemical features of HG base-pairs in sequence-specific DNA recognition, and therefore, provided evidence for a functional role for HG base-pairs *in vivo*.

The first crystallographic observation of HG base-pairs in a protein-DNA complex was reported by Rice *et al.* in 1996<sup>67</sup> who visualized a single A-T HG base-pair immediately adjacent to a nicked site in the X-ray structure of a highly bent (>160 degrees) 35 base-pair (bp) DNA bound to the integration host factor (IHF) protein (Figure 4). Interestingly, a hydrogen bond was observed between the backbone amide group of an arginine residue and N3 of the *syn* A, suggesting specific recognition of the Watson-Crick face in the HG base-pair. However, the nick is involved in crystal packing with a neighboring molecule in the complex and HG formation helps move the phosphate backbone away from a neighboring molecule. In addition, the protein makes specific contacts with N3 of an *anti* A in a symmetric site in the DNA lacking the nick, suggesting interactions that are specific for WC rather than HG base-pairing. Moreover, NMR studies of IHF binding to a shorter recognition sequence containing the first nicked site argue against the presence of an A-T HG base-pair in solution.<sup>68</sup>

Subsequent X-ray structures of TATA elements bound to the TATA box-binding protein (TBP) revealed a G-C<sup>+</sup> HG base-pair in the mutant TATAAAC box in a region of DNA unwinding and intercalation.<sup>69</sup> No direct contacts were observed between the *syn* guanine base and the protein. However, the HG base-pair appears to contribute to binding by preventing steric clashes between the protein leucine 72 and the guanine exocyclic NH<sub>2</sub>, while still preserving favorable van der Waals contacts with two neighboring phenylalanine residues. A second G-C HG base-pair was observed but attributed to crystal packing forces. Interestingly, the ~150-fold weaker binding affinity observed for TBP to this mutant TATA box,<sup>70</sup> which could be correlated to the selection of a transient HG over a WC base-pair at that site,<sup>71</sup> has been implicated in the transcriptional regulation of the human osteocalcin gene.<sup>72</sup> This observation suggests a biological role for the formation of a G-C<sup>+</sup> HG base-pair at the mutant promoter site.

Both IHF and TBP induce large distortions in the DNA, which could facilitate formation of HG base-pairs. In contrast, Wolberger and colleagues observed a single A-T HG base-pair within an otherwise undistorted B-form WC duplex in the X-ray structure of MAT 2 homeodomain non-specifically bound to DNA. Ya Van der Waals contacts were observed between an arginine side chain and the *syn* adenine base as well as the sugar-phosphate backbone of the adenine and the neighboring thymine. Once again, the HG base-pair appears to avoid unfavorable steric clashes that would otherwise arise with a WC base-pair. The HG

base-pair is accommodated within the duplex DNA without inducing major distortions, even for the directly neighboring base-pairs. The ease with which HG base-pairs could seamlessly fit within B-DNA raised the possibility that HG base-pairs may have been incorrectly assigned to be WC base-pairs due to misinterpretation of ambiguous electron density at medium to low resolution.<sup>73</sup>

Two neighboring A-T HG base-pairs were subsequently observed in structures of a palindromic CATG/CATG sequence bound to the DNA binding domain of p53.<sup>74</sup> Although no direct contacts are observed with the *syn* adenines, the formation of the HG base-pairs results in a narrowed minor groove in the region flanking the CATG site, leading to enhanced negative electrostatic potential that is further stabilized by insertion of the positively charged arginine side chains. Remarkably, these HG base-pairs adopt WC conformation in X-ray structures with a longer spacer length<sup>74</sup> or a different intervening sequence<sup>75,76</sup> between DNA half-sites, which is accompanied by a different organization between p53 dimers, altered DNA helix conformation, and that also yield different DNA-tetramer binding affinities.<sup>77</sup> These studies suggest that WC and HG base-pairs likely exist in equilibrium with each other and that their selection in DNA-p53 complexes is largely dictated by the nature of the DNA binding sequence.

## **Damaged DNA**

By the 1960s, it had become clear that DNA could be damaged by exogenous and endogenous factors, and that this in turn may be linked to disease states such as cancer. During the 1970s and 1980s, enzymes that recognize and repair damaged DNA began to be uncovered, resulting in great interest in characterizing the structure of damaged DNA. These studies showed that HG base-pairing provides an important mechanism for stacking and hydrogen bonding, in cases where the Watson-Crick face of the purine bases is damaged, preventing favorable WC base-pairing.

The first evidence for HG-type base-pairs in damaged DNA was reported in the late 1980s in solution NMR studies by Patel and co-workers showing that guanine adducts on the Watson-Crick edge or the C8 positions strongly favor a *syn* base orientation. 80–83 Subsequent NMR studies showed HG-type pairing in various purine lesions, including WC face alkylation adducts (e.g. 1,N2-propanoguanine<sup>84,85</sup> and 1,N2-ethenoguanine<sup>86</sup>), the bulky guanine C8 mutagenic adduct aminofluorine-C8-guanine<sup>83,87</sup>, and the common mutagenic lesion N1-methyladenine (Figure 5). 88 The direct observation of HG base-pairing (rather than extrahelical states) in a wide variety of lesions in naked DNA in the 1990s and 2000s established HG base-pairs as an energetically closer alternative to WC base pairs.

There is great speculation and experimental evidence that HG-type pairs play important roles in DNA damage and mismatch repair. For example, it is likely that the enzyme AlkB, which repairs the mutagenic lesion N1-methyladenine, initially recognizes the HG base-pair between N1-methyladenine and thymine (Figure 5)<sup>88,89</sup> before flipping out the damaged purine for oxidative demethylation. The flipping of one purine base to a *syn* conformation is also often observed in purine-purine mismatches, where the *syn-anti* base-pair configuration affords a shorter helical radius that can be more readily accommodated within B-DNA as compared to the *anti-anti* configuration. There is X-ray structural evidence that the DNA mismatch repair enzyme MutS specifically recognizes HG type purine-purine and purine-pyrimidine mismatches, even though they may not be the dominant conformation in unbound DNA, by making specific hydrophobic and hydrogen bonding minor groove contacts with the *syn* adenine/guanine base in A-C, A-A, and G-G mismatches. <sup>90</sup> The recognition of the increased population of *syn-anti* rather than *anti-anti* configuration in certain mismatched base-pairs may help the enzyme discriminate against undamaged *anti-*

anti Watson-Crick base-pairs. Thus, HG base-pairs not only provide a mechanism for maintaining the overall structural integrity of damaged or incorrectly replicated DNA, they can play an important role in DNA repair mechanisms.

It is worth noting that HG base-pairs have also been observed in DNA containing non-natural modifications in the sugar-phosphate backbone, including the addition of an ethylene bridge between C3 and C5 in bicyclo-DNA, which fixes the gamma backbone torsion angle to a non-canonical orientation, <sup>91</sup> a single-residue substitution of sugar O4 with a methylene group, <sup>92</sup> or in dinucleotide d(TA) analogs containing a nonionic diisopropylsilyl-modified backbone at very low temperatures. <sup>93</sup>

### **DNA Replication**

Watson-Crick base-pairs were the most important aspect of the DNA double helix structure because, as succinctly stated in the very last sentence of their 1953 Nature paper, it "immediately suggests a possible copying mechanism for the genetic material." Four years later, Kornberg discovered the enzyme that catalyzed template DNA replication<sup>94</sup> and ensuing biochemical and structural studies established that DNA polymerases replicate DNA by Watson-Crick pairing dNTP with the template strand.

During the 1990s, studies revealed that certain families of DNA polymerases (the X and Y families) contributed to damage-induced mutagenesis. It was later shown that some members of the Y family DNA polymerases efficiently bypass DNA damage by replicating the template DNA via HG rather than WC base-pairing. HG-based replication was first visualized in X-ray structures of an archaeal DNA Pol homolog, Dpo4, by Yang and colleagues nearly a decade ago. 95 The structure showed that Dpo4 replicates UV cross-linked thymine dimers by forming a HG base-pair between the 5 thymine and an incoming ddATP, thus avoiding backbone distortion and allowing discrimination against guanine and pyrimidines. 95

Aggarwal and co-workers subsequently showed using X-ray crystallography and biochemical experiments that another member of this family, human DNA Pol, employs HG base-paring as a general mechanism to replicate both damaged and undamaged DNA. HG base-paring as a general mechanism to replicate both damaged and undamaged DNA. A striking X-ray structure of Pol showed a template adenine in the active site of the enzyme adopting a *syn* conformation and forming a HG base-pair with an incoming dTTP (Figure 6). HG ability to insert the correct nucleotide across an adenine base also provided a rationale for prior biochemical studies showing a much higher efficiency of correct base incorporation across a templating adenine than across a templating thymine, which in fact favors G misincorporation because of its high propensity for forming of *anti*-G-T wobble base-pair. This raised HG base-pairs to a prominent position reserved previously only for WC base-pairs; they provided a basis for copying DNA.

The proposal that hPol replicates DNA via HG base-pairing was quickly met with skepticism. In an accompanying *News and Views* article, Wang<sup>98</sup> pointed out that, based on the weak electron density for the active site A-T base-pair, it is difficult to resolve a WC from a HG conformation. He also questioned the ability of such a polymerase to form protonated G-C<sup>+</sup> base-pairs at physiological pH, given the low intrinsic p $K_a$  of cytosine N3 (~ 4.2 – 4.4). <sup>99</sup> Aggarwal *et al.* later put the matter to rest by (i) solving X-ray structures of Pol , unambiguously showing a protonated G-C<sup>+</sup> HG base-pair at pH 6.5, reinforcing their hypothesis that Pol has evolved to favor HG base-pairing by constraining the backbone C1 -C1 distance between template and incoming nucleotide<sup>100</sup> and (ii) showing selective inhibition of DNA synthesis by Pol but not other polymerases when using 7-deazaadenine or 7-deazaguanine, which are incapable of forming HG base-pairing, as the templating residue. <sup>101</sup> Several other structures capturing DNA synthesis by Pol followed, ultimately

demonstrating that major purine alkylation and oxidation lesions, including 1,N6-ethenoadenine, <sup>102</sup> N2-ethylguanine, <sup>103</sup> O6-methylguanine <sup>104</sup> and 8-oxoguanine, <sup>105</sup> adopted a *syn* conformation and, where possible, formed HG type base-pairs with incoming complementary pyrimidine and purine nucleotides (Figure 6) (reviewed in Makarova *et al.* <sup>106</sup>). These observations, in conjunction with biological studies showing that Pol was important for cell survival in the presence of alkylating agents <sup>107</sup>, <sup>108</sup> and oxidative stress, <sup>109</sup> provide the most compelling evidence to date for a biological function for HG base-pairs in duplex DNA.

### **Transient Hoogsteen Base-Pairs**

The earliest fiber X-ray diffraction studies of DNA highlighted its polymorphic nature and the ability of the double helix to adopt different forms depending on environmental conditions and sequence contexts. Subsequent studies showed that DNA does indeed come in many different forms and that even B-DNA is not rigid, but rather, can undergo large deformations and thermal fluctuations in a sequence-dependent, biologically important manner. <sup>110</sup> This flexibility was not confined to the weakly constrained sugar and phosphodiester backbone, but also includes the Watson-Crick base-pairs themselves. <sup>111</sup> Chemical probing and hydrogen exchange studies spanning the 1970s–1990s established that WC base-pairs break apart and open at millisecond timescales and that the open state exists in at most ~0.002% abundance for A-T or ~0.00008% for G-C base-pairs. <sup>112–116</sup> There are now several X-ray structures that capture these open states of the base-pairs when bound to proteins that establish their functional significance.

Two years ago, NMR studies from our laboratory showed that both A-T and G-C Watson-Crick base-pairs can transiently undergo excursions toward HG base-pairs in duplex DNA.<sup>71,117</sup> The transient HG base-pairs were characterized with the use of recently developed NMR R<sub>1</sub> relaxation dispersion spectroscopic methods that make it possible to observe and structurally characterize fleeting states of macromolecules. 118-120 The transient HG base-pairs had populations of  $\sim 0.1 - 1\%$ , making them nearly three orders of magnitude more abundant than the open state, with the G-C<sup>+</sup> HG base-pairs being less abundant than their A-T counterparts at physiological pH by at least a factor of 20 due to an additional required protonation event at cytosine N3.<sup>17</sup> The transient HG base-pairs have lifetimes on the order of hundreds of microseconds to milliseconds (~ 0.3 to 1.5 ms), which are significantly longer than the lifetimes of base-pair open states found to be in the nanosecond range. 116 The free energy and enthalpy of the WC-to-HG transition were found to closely match those of base-pair opening, 116 suggesting that the transition may be limited by a basepair breaking event that could be coupled to the purine anti-syn isomerization, inside or outside of the double helix. It is remarkable that the HG base-pairs are energetically less favorable than WC counterparts by a mere ~3 kcal/mol in the case of A-T base-pairs, roughly the equivalent of one strong hydrogen-bond. These energetic differences are small compared to forces that exist in cells due to protein interactions, torsional stress due to binding and supercoiling, or those applied due to crystal packing forces or that arise from changing pH. Studies suggest that the transient HG base-pairs occur universally across all DNA sequence contexts, in a non-cooperative manner, and with small, albeit significant, sequence-specific differences in population and lifetimes (unpublished results).

The picture that emerges is one in which every base-pair in DNA exists as an equilibrium of WC and HG base-pairs, with external parameters operating on the DNA resolving one or the other base-pair type. This helps explain the long and controversial observation of WC versus HG – small changes in conditions can favor one form over the other. It is striking that the difference in the abundance of transient G-C<sup>+</sup> and A-T HG base-pairs mirrors the differences in efficiency observed in Pol replication of A/T versus G/C. The HG base-pairs

transiently expose the Watson-Crick faces of purines, and may potentially help explain the much greater abundance of N1 methylation in adenine versus guanine. Most importantly, the observation of transient HG base-pairs in duplex DNA, with comparable energetics to WC, raises the possibility that HG base-pairs exist in much greater abundance *in vivo*, particularly in A-T rich regions of the genome. When combined with the current difficulties in resolving WC from HG based on X-ray diffraction data, it may well be the case that there are more HG base-pairs in X-ray structures currently deposited in the PDB that have gone undetected, particularly for A-T base-pairs. We hope that this review provides the impetus to be more critical of the interpretation of X-ray diffraction data to rule out the possibility of HG base-pairs.

#### **Future outlook**

Thus, 60 years later, there are fundamental questions that remain to be answered regarding the structure of the DNA double helix.

- 1. What fraction of the existing DNA structures in the Protein Data Bank contain base-pairs that, due to biases and poor electron density, have been misinterpreted to be WC rather than HG? It should be relatively straightforward to re-interpret the electron density for DNA structures and to examine, base-pair by base-pair, to what extent the data favor WC versus HG. We predict that there will be many structures that will be revised to include uncertain base-pair conformations, or even HG base-pairs, particularly for A-T base-pairs.
- 2. Are HG base-pairs that are observed in X-ray structures also observed under solution conditions? Here, there is a need to develop new methods to allow the characterization of HG versus WC under solution conditions. Developments in NMR that can allow studies of large protein-DNA complexes will undoubtedly be important –but other, high-throughput, approaches are also needed to streamline such applications.
- 3. If HG base-pairs do occur, to what extent are they functionally important? Are the HG base-pairs 'passively' present because they merely provide a more stable form under a particular condition or do they 'actively' participate in biological function? Here, there is a danger of rushing to interpret results with 7-deazapurines, which diminish the ability to form HG base-pairs. It is conceivable that the 7-deazapurines can still form distorted HG base-pairs that preserve function, or affect an aspect of function *in vivo* that is not explored in such *in vitro* studies.
- To what extent do HG or any other type of base-pair occur in vivo? Today, there are no methods for characterizing, at the atomic level, the nature of the base-pairs that hold together DNA duplexes in vivo. In Eukaryotes, DNA is wrapped around nucleosomes and packaged with proteins to form chromatin fibers that make up chromosomes. The DNA is subjected to extreme packing and supercoiling, which present forces that in all likelihood exceed the energetic differences between WC and HG observed in duplex DNA. Given the growing evidence that HG base-pairs tend to be favored under tight packing conditions, and in regions of stress, we can predict that the genome may in fact be enriched with HG base-pairs in vivo relative to relaxed duplex DNA. One can imagine the existence of 'HG islands' in A-T rich regions that have been shown to be important for DNA minor groove recognition by a variety of protein factors and anticancer drugs regulating replication/ transcription, DNA bending, supercoiling induced DNA destabilization, nucleosome positioning, and chromosomal translocation. 121-128 A challenge for the future will be the development of methods for visualizing the high resolution structure of DNA in vivo.

### **Acknowledgments**

This research was supported by an NIH grant (R01GM089846) and Agilent Foundation. E.N.N. is a Robert Black Fellow of the Damon Runyon Cancer Research Foundation (DRG-2159-13).

#### References

- 1. Watson JD, Crick FH. Nature. 1953; 171:737–738. [PubMed: 13054692]
- 2. Watson JD, Crick FH. Nature. 1953; 171:964-967. [PubMed: 13063483]
- 3. Franklin RE, Gosling RG. Nature. 1953; 171:740–741. [PubMed: 13054694]
- 4. Franklin RE, Gosling RG. Acta Cryst. 1953; 6:673.
- Wing R, Drew H, Takano T, Broka C, Tanaka S, Itakura K, Dickerson RE. Nature. 1980; 287:755–758. [PubMed: 7432492]
- 6. Dickerson RE, Drew HR. J Mol Biol. 1981; 149:761-786. [PubMed: 6273591]
- 7. Hoogsteen K. Acta Crystallogr. 1959; 12:822–823.
- 8. Hoogsteen K. Acta crystallographica Section A, Foundations of crystallography. 1963; 16:907.
- 9. Wang AH, Quigley GJ, Kolpak FJ, Crawford JL, van Boom JH, van der Marel G, Rich A. Nature. 1979; 282:680–686. [PubMed: 514347]
- 10. Felsenfeld G, Rich A. Biochim Biophys Acta. 1957; 26:457–468. [PubMed: 13499402]
- 11. Sobell HM, Tomita KI, Rich A. Proc Natl Acad Sci U S A. 1963; 49:885–892. [PubMed: 13989773]
- 12. Corey RB, Pauling L. Arch Biochem Biophys. 1956; 65:164–181. [PubMed: 13373416]
- 13. Haschemeyer AE, Sobell HM. Acta Crystallogr. 1965; 19:125–130. [PubMed: 5896866]
- 14. Haschemeyer AE, Sobell HM. Proc Natl Acad Sci U S A. 1963; 50:872-877. [PubMed: 14082351]
- 15. Mathews FS, Rich A. Journal of molecular biology. 1964; 8:89-95. [PubMed: 14149966]
- 16. Courtois Y, Fromageot P, Guschlbauer W. Eur J Biochem. 1968; 6:493-501. [PubMed: 5701966]
- 17. Nikolova EN, Goh GB, Brooks CL 3rd, Al-Hashimi HM. J Am Chem Soc. 2013
- 18. Haas BL, Guschlbauer W. Nucleic Acids Res. 1976; 3:205-218. [PubMed: 1250698]
- 19. Kallenbach NR, Daniel WE Jr, Kaminker MA. Biochemistry. 1976; 15:1218–1224. [PubMed: 1252442]
- 20. Quigley GJ, Ughetto G, van der Marel GA, van Boom JH, Wang AH, Rich A. Science. 1986; 232:1255–1258. [PubMed: 3704650]
- 21. Pauling L, Corey RB. Proc Natl Acad Sci U S A. 1953; 39:84–97. [PubMed: 16578429]
- 22. Kearns DR, Patel DJ, Shulman RG. Nature. 1971; 229:338–339. [PubMed: 4927207]
- 23. Cross AD, Crothers DM. Biochemistry. 1971; 10:4015–4023. [PubMed: 5161029]
- 24. Wong YP, Kearns DR, Reid BR, Shulman RG. Journal of Molecular Biology. 1972; 72:725–740. [PubMed: 4573845]
- 25. Shulman RG, Hilbers CW, Wong YP, Wong KL, Lightfoot DR, Reid BR, Kearns DR. Proc Natl Acad Sci U S A. 1973; 70:2042–2045. [PubMed: 4579011]
- Crothers DM, Hilbers CW, Shulman RG. Proc Natl Acad Sci U S A. 1973; 70:2899–2901.
   [PubMed: 4517943]
- 27. Patel DJ, Tonelli AE. Biopolymers. 1974; 13:1943–1964. [PubMed: 4433696]
- 28. Day RO, Seeman NC, Rosenberg JM, Rich A. Proc Natl Acad Sci U S A. 1973; 70:849–853. [PubMed: 4514996]
- Rosenberg JM, Seeman NC, Kim JJ, Suddath FL, Nicholas HB, Rich A. Nature. 1973; 243:150– 154. [PubMed: 4706285]
- 30. Rich A. Q Rev Biophys. 2009; 42:117–137. [PubMed: 19638248]
- 31. Arentzen R, Van Boeckel C, Van der Marel G, Van Boom J. 1979
- 32. de Rooij JFM, Wille-Hazeleger G, van Deursen PH, Serdijn J, van Boom JH. Recueil des Travaux Chimiques des Pays-Bas. 1979; 98:537–548.
- 33. Mitsui Y, Langridge R, Shortle BE, Cantor CR, Grant RC, Kodama M, Wells RD. Nature. 1970; 228:1166–1169. [PubMed: 4321098]

- 34. Pohl FM, Jovin TM. Journal of Molecular Biology. 1972; 67:375–396. [PubMed: 5045303]
- 35. Drew HR, Dickerson RE. Embo J. 1982; 1:663–667. [PubMed: 7188355]
- 36. Davies DR, Baldwin RL. Journal of Molecular Biology. 1963; 6:251–255. [PubMed: 14025400]
- 37. Selsing E, Arnott S. Journal of Molecular Biology. 1975; 98:243–248. [PubMed: 1195382]
- 38. Ikehara M, Hattori M, Fukui T. Eur J Biochem. 1972; 31:329–334. [PubMed: 4647181]
- 39. Ishikawa F, Frazier J, Howard FB, Miles HT. Journal of Molecular Biology. 1972; 70:475–490. [PubMed: 5083147]
- 40. Hattori M, Ikehara M, Miles HT. Biochemistry. 1974; 13:2754–2761. [PubMed: 4847543]
- 41. Liu K, Miles HT, Frazier J, Sasisekharan V. Biochemistry. 1993; 32:11802–11809. [PubMed: 8218251]
- 42. Segers-Nolten GM, Sijtsema NM, Otto C. Biochemistry. 1997; 36:13241–13247. [PubMed: 9341213]
- 43. Blommers MJ, van de Ven FJ, van der Marel GA, van Boom JH, Hilbers CW. Eur J Biochem. 1991; 201:33–51. [PubMed: 1915376]
- 44. Ronning DR, Guynet C, Ton-Hoang B, Perez ZN, Ghirlando R, Chandler M, Dyda F. Mol Cell. 2005; 20:143–154. [PubMed: 16209952]
- 45. Abrescia NG, Thompson A, Huynh-Dinh T, Subirana JA. Proc Natl Acad Sci U S A. 2002; 99:2806–2811. [PubMed: 11880632]
- 46. Abrescia NG, Gonzalez C, Gouyette C, Subirana JA. Biochemistry. 2004; 43:4092–4100. [PubMed: 15065851]
- 47. Pous J, Urpi L, Subirana JA, Gouyette C, Navaza J, Campos JL. J Am Chem Soc. 2008; 130:6755–6760. [PubMed: 18447354]
- 48. De Luchi D, Tereshko V, Gouyette C, Subirana JA. Chembiochem. 2006; 7:585–587. [PubMed: 16485316]
- 49. Wang AH, Ughetto G, Quigley GJ, Hakoshima T, van der Marel GA, van Boom JH, Rich A. Science. 1984; 225:1115–1121. [PubMed: 6474168]
- 50. Ward DC, Reich E, Goldberg IH. Science. 1965; 149:1259–1263. [PubMed: 5318292]
- 51. Sato K, Shiratori O, Katagiri K. J Antibiot (Tokyo). 1967; 20:270-276. [PubMed: 4170544]
- 52. Ughetto G, Wang AH, Quigley GJ, van der Marel GA, van Boom JH, Rich A. Nucleic Acids Res. 1985; 13:2305–2323. [PubMed: 4000957]
- 53. Mendel D, Dervan PB. Proc Natl Acad Sci U S A. 1987; 84:910–914. [PubMed: 3103127]
- 54. Herr W. Proc Natl Acad Sci U S A. 1985; 82:8009-8013. [PubMed: 3865212]
- 55. Scholten PM, Nordheim A. Nucleic Acids Res. 1986; 14:3981–3993. [PubMed: 3714469]
- 56. McLean MJ, Waring MJ. J Mol Recognit. 1988; 1:138–151. [PubMed: 3273225]
- 57. Portugal J, Fox KR, McLean MJ, Richenberg JL, Waring MJ. Nucleic Acids Res. 1988; 16:3655–3670. [PubMed: 3287335]
- 58. McLean MJ, Seela F, Waring MJ. Proc Natl Acad Sci U S A. 1989; 86:9687–9691. [PubMed: 2602369]
- 59. Sayers EW, Waring MJ. Biochemistry. 1993; 32:9094–9107. [PubMed: 8369281]
- 60. Gao XL, Patel DJ. Biochemistry. 1988; 27:1744-1751. [PubMed: 3365421]
- 61. Gilbert DE, van der Marel GA, van Boom JH, Feigon J. Proc Natl Acad Sci U S A. 1989; 86:3006–3010. [PubMed: 2717609]
- 62. Gilbert DE, Feigon J. Biochemistry. 1991; 30:2483–2494. [PubMed: 2001374]
- 63. Gilbert DE, Feigon J. Nucleic Acids Res. 1992; 20:2411–2420. [PubMed: 1598199]
- 64. Park JY, Choi BS. J Biochem. 1995; 118:989–995. [PubMed: 8749317]
- 65. Gallego J, Luque FJ, Orozco M, Burgos C, Alvarez-Builla J, Rodrigo MM, Gago F. J Med Chem. 1994; 37:1602–1609. [PubMed: 8201593]
- Gallego J, Luque FJ, Orozco M, Gago F. J Biomol Struct Dyn. 1994; 12:111–129. [PubMed: 7848562]
- 67. Rice PA, Yang S, Mizuuchi K, Nash HA. Cell. 1996; 87:1295–1306. [PubMed: 8980235]

68. Dhavan GM, Lapham J, Yang S, Crothers DM. Journal of Molecular Biology. 1999; 288:659–671. [PubMed: 10329171]

- Patikoglou GA, Kim JL, Sun L, Yang SH, Kodadek T, Burley SK. Genes Dev. 1999; 13:3217–3230. [PubMed: 10617571]
- 70. Hoopes BC, LeBlanc JF, Hawley DK. J Mol Biol. 1998; 277:1015–1031. [PubMed: 9571019]
- 71. Nikolova EN, Kim E, Wise AA, O'Brien PJ, Andricioaei I, Al-Hashimi HM. Nature. 2011; 470:498–502. [PubMed: 21270796]
- 72. Meyer T, Gustafsson JA, Carlstedt-Duke J. DNA Cell Biol. 1997; 16:919–927. [PubMed: 9303434]
- 73. Aishima J, Gitti RK, Noah JE, Gan HH, Schlick T, Wolberger C. Nucleic Acids Res. 2002; 30:5244–5252. [PubMed: 12466549]
- 74. Kitayner M, Rozenberg H, Rohs R, Suad O, Rabinovich D, Honig B, Shakked Z. Nat Struct Mol Biol. 2010; 17:423–429. [PubMed: 20364130]
- 75. Malecka KA, Ho WC, Marmorstein R. Oncogene. 2009; 28:325–333. [PubMed: 18978813]
- 76. Chen Y, Dey R, Chen L. Structure. 2010; 18:246-256. [PubMed: 20159469]
- 77. Kitayner M, Rozenberg H, Kessler N, Rabinovich D, Shaulov L, Haran TE, Shakked Z. Mol Cell. 2006; 22:741–753. [PubMed: 16793544]
- 78. Alexander P. Advanc Cancer Res. 1954:2.1.
- 79. Friedberg EC. Cell Res. 2008; 18:3-7. [PubMed: 18157159]
- 80. Patel DJ, Shapiro L, Kozlowski SA, Gaffney BL, Jones RA. Journal of Molecular Biology. 1986; 188:677–692. [PubMed: 3016288]
- 81. Kalnik MW, Li BF, Swann PF, Patel DJ. Biochemistry. 1989; 28:6182–6192. [PubMed: 2789994]
- 82. Kouchakdjian M, Marinelli E, Gao XL, Johnson F, Grollman A, Patel D. Biochemistry. 1989; 28:5647–5657. [PubMed: 2775729]
- 83. Norman D, Abuaf P, Hingerty BE, Live D, Grunberger D, Broyde S, Patel DJ. Biochemistry. 1989; 28:7462–7476. [PubMed: 2819081]
- 84. Singh US, Moe JG, Reddy GR, Weisenseel JP, Marnett LJ, Stone MP. Chem Res Toxicol. 1993; 6:825–836. [PubMed: 8117922]
- 85. Weisenseel JP, Reddy GR, Marnett LJ, Stone MP. Chem Res Toxicol. 2002; 15:127–139. [PubMed: 11849038]
- 86. Shanmugam G, Kozekov ID, Guengerich FP, Rizzo CJ, Stone MP. Chem Res Toxicol. 2008; 21:1795–1805. [PubMed: 18693701]
- 87. Mao B, Hingerty BE, Broyde S, Patel DJ. Biochemistry. 1998; 37:81–94. [PubMed: 9425028]
- 88. Yang H, Zhan Y, Fenn D, Chi LM, Lam SL. FEBS Lett. 2008; 582:1629–1633. [PubMed: 18435925]
- 89. Lu L, Yi C, Jian X, Zheng G, He C. Nucleic Acids Res. 2010; 38:4415–4425. [PubMed: 20223766]
- 90. Natrajan G, Lamers MH, Enzlin JH, Winterwerp HH, Perrakis A, Sixma TK. Nucleic Acids Res. 2003; 31:4814–4821. [PubMed: 12907723]
- 91. Bolli M, Litten JC, Schutz R, Leumann CJ. Chem Biol. 1996; 3:197–206. [PubMed: 8807846]
- 92. Isaksson J, Zamaratski E, Maltseva TV, Agback P, Kumar A, Chattopadhyaya J. J Biomol Struct Dyn. 2001; 18:783–806. [PubMed: 11444368]
- 93. Xiao Z, Weisz K. J Am Chem Soc. 2010; 132:3862-3869. [PubMed: 20180555]
- Lehman IR, Bessman MJ, Simms ES, Kornberg A. J Biol Chem. 1958; 233:163–170. [PubMed: 13563462]
- 95. Ling H, Boudsocq F, Plosky BS, Woodgate R, Yang W. Nature. 2003; 424:1083–1087. [PubMed: 12904819]
- Nair DT, Johnson RE, Prakash S, Prakash L, Aggarwal AK. Nature. 2004; 430:377–380. [PubMed: 15254543]
- 97. Tissier A, McDonald JP, Frank EG, Woodgate R. Genes & development. 2000; 14:1642–1650. [PubMed: 10887158]
- 98. Wang J. Nature. 2005; 437:E6-7. discussion E7. [PubMed: 16163299]

99. Izatt RM, Christensen JJ, Rytting JH. Chem Rev. 1971; 71:439-481. [PubMed: 5126179]

- 100. Nair DT, Johnson RE, Prakash L, Prakash S, Aggarwal AK. Structure. 2005; 13:1569–1577. [PubMed: 16216587]
- 101. Johnson RE, Prakash L, Prakash S. Proc Natl Acad Sci U S A. 2005; 102:10466–10471.
  [PubMed: 16014707]
- Nair DT, Johnson RE, Prakash L, Prakash S, Aggarwal AK. Nature structural & molecular biology. 2006; 13:619–625.
- Pence MG, Blans P, Zink CN, Hollis T, Fishbein JC, Perrino FW. J Biol Chem. 2009; 284:1732–1740. [PubMed: 18984581]
- 104. Pence MG, Choi JY, Egli M, Guengerich FP. J Biol Chem. 2010; 285:40666–40672. [PubMed: 20961860]
- 105. Kirouac KN, Ling H. Proc Natl Acad Sci U S A. 2011; 108:3210–3215. [PubMed: 21300901]
- 106. Makarova AV, Kulbachinskiy AV. Biochemistry (Mosc). 2012; 77:547–561. [PubMed: 22817454]
- Johnson RE, Yu SL, Prakash S, Prakash L. Mol Cell Biol. 2007; 27:7198–7205. [PubMed: 17698580]
- 108. Plosky BS, Frank EG, Berry DA, Vennall GP, McDonald JP, Woodgate R. Nucleic Acids Res. 2008; 36:2152–2162. [PubMed: 18281311]
- Petta TB, Nakajima S, Zlatanou A, Despras E, Couve-Privat S, Ishchenko A, Sarasin A, Yasui A, Kannouche P. Embo J. 2008; 27:2883–2895. [PubMed: 18923427]
- 110. Rich A. Gene. 1993; 135:99-109. [PubMed: 8276285]
- 111. Travers AA. Philos Transact A Math Phys Eng Sci. 2004; 362:1423–1438.
- 112. Englander SW, Downer NW, Teitelbaum H. Annu Rev Biochem. 1972; 41:903–924. [PubMed: 4563445]
- 113. Gueron M, Kochoyan M, Leroy JL. Nature. 1987; 328:89–92. [PubMed: 3037381]
- 114. Gueron M, Leroy JL. Methods in enzymology. 1995; 261:383–413. [PubMed: 8569504]
- 115. Russu IM. Methods in enzymology. 2004; 379:152–175. [PubMed: 15051357]
- 116. Coman D, Russu IM. J Biol Chem. 2005; 280:20216-20221. [PubMed: 15784615]
- 117. Nikolova EN, Gottardo FL, Al-Hashimi HM. J Am Chem Soc. 2012; 134:3667–3670. [PubMed: 22309937]
- 118. Massi F, Johnson E, Wang C, Rance M, Palmer AG 3rd. J Am Chem Soc. 2004; 126:2247–2256. [PubMed: 14971961]
- Korzhnev DM, Orekhov VY, Kay LE. J Am Chem Soc. 2005; 127:713–721. [PubMed: 15643897]
- 120. Hansen AL, Nikolova EN, Casiano-Negroni A, Al-Hashimi HM. J Am Chem Soc. 2009; 131:3818–3819. [PubMed: 19243182]
- 121. Bowater R, Aboul-ela F, Lilley DM. Biochemistry. 1991; 30:11495-11506. [PubMed: 1747368]
- 122. Carrera P, Azorin F. Nucleic Acids Res. 1994; 22:3671-3680. [PubMed: 7937076]
- 123. Woynarowski JM, Trevino AV, Rodriguez KA, Hardies SC, Benham CJ. J Biol Chem. 2001; 276:40555–40566. [PubMed: 11487576]
- 124. Rohs R, West SM, Sosinsky A, Liu P, Mann RS, Honig B. Nature. 2009; 461:1248–1253. [PubMed: 19865164]
- 125. Segal E, Widom J. Curr Opin Struct Biol. 2009; 19:65–71. [PubMed: 19208466]
- 126. Gordon BR, Li Y, Cote A, Weirauch MT, Ding P, Hughes TR, Navarre WW, Xia B, Liu J. Proc Natl Acad Sci U S A. 2011; 108:10690–10695. [PubMed: 21673140]
- 127. Kato T, Kurahashi H, Emanuel BS. Curr Opin Genet Dev. 2012; 22:221–228. [PubMed: 22402448]
- 128. Rajewska M, Wegrzyn K, Konieczny I. FEMS Microbiol Rev. 2012; 36:408–434. [PubMed: 22092310]
- 129. Zheng G, Lu XJ, Olson WK. Nucleic Acids Res. 2009; 37:W240-246. [PubMed: 19474339]

**Figure 1.** Chemical structures of A-T and G-C Watson-Crick (WC) and Hoogsteen (HG) base-pairs.

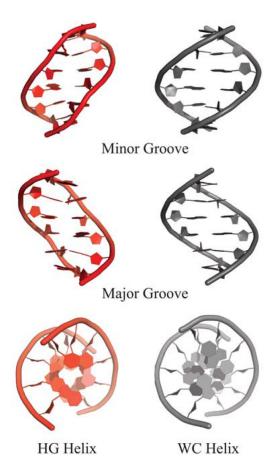


Figure 2. Comparison between an ideal Watson-Crick (WC) and a Hoogsteen (HG) double helix from the crystal structure of  $d(AT)_3$  (PDB ID: 1RSB). The B-DNA helix was built using w3DNA. 129

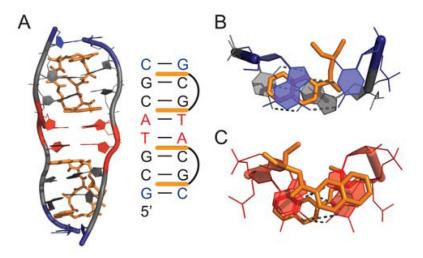
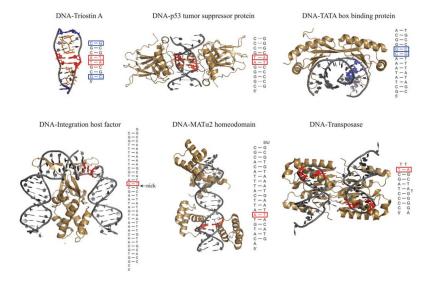


Figure 3.
Hoogsteen (HG) base-pairs in a DNA-antibiotic complex. (A) 3D (left) and secondary (right) structures of a DNA octamer in complex with the bis-intercalating antibiotic triostin A (orange), highlighting internal A-T HG base-pairs (red) and terminal G-C<sup>+</sup> base-pairs (blue) (PDB ID: 1VS2). Top view of the (B) terminal G-C<sup>+</sup>HG base-pair and (C) helical A-T HG base-pairs depicting characteristic hydrogen bonds (dashed lines) and stacking between the two base-pairs and the quinoxaline moiety of triostin A.



**Figure 4.**Hoogsteen (HG) base-pairs in DNA-drug/protein complexes. 3D structure (left) and DNA sequence (right) of DNA complexes with triostin A (PDB ID: 1VS2), p53 tumor suppressor protein (PDB ID: 3IGL), TATA-box binding protein (TBP) (PDB ID: 1QN3), integration host factor (IHF) (PDBID: 1IHF), MAT 2 homeodomain (PDB ID: 1K61), and transposase (PDB ID: 2A6O), highlighting the A-T (red) and G-C<sup>+</sup> (blue) HG base-pairs (5IU, 5-iodouridine; dashed lines indicate unpaired residues).

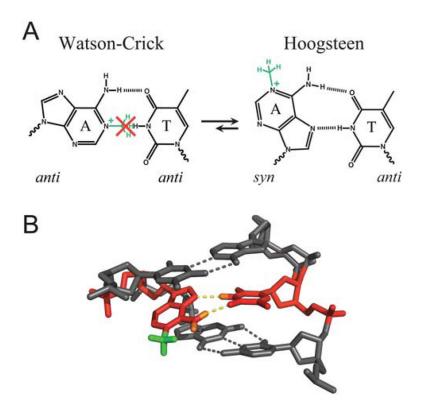


Figure 5. The N1-methyladenine (N1-Me-dA) damage favors Hoogsteen base-pairing with T. A) The methyl damage (green) on A N1 causes a steric clash in the WC base-pair, which favors formation of the alternative HG base-pair that relieves the steric clash. B) Crystal structure of an HG base-pair between N1-Me-dA and T where the methyl modification is shown in green (PDB ID: 3H8O).

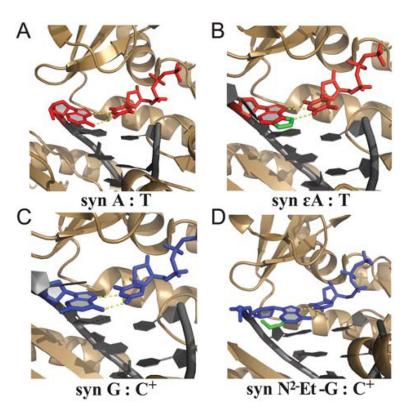


Figure 6.
DNA replication by human lesion bypass DNA polymerase (hPol) proceeds via Hoogsteen (HG) base-pairing. Shown are complexes between hPol and an HG base-pair between (A) a template A and an incoming dTTP (PDB ID: 1T3N), (B) a template 1,N6-ethenoadenine (A) and an incoming dTTP (PDB ID: 2DPJ), (C) a template G and an incoming dCTP (PDB ID: 2ALZ), and (D) a template N2-ethylguanine (N2-Et-G) and an incoming dCTP (PDB ID: 3EPG). The damaged sites in A and N2-Et-G are highlighted in green.

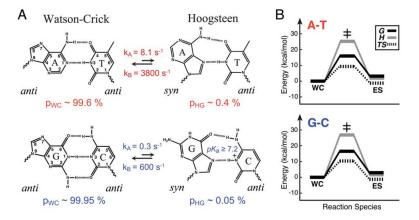


Figure 7. The conformational exchange between a ground state Watson-Crick (WC) and a transient Hoogsteen (HG) base-pair can be monitored using NMR relaxation dispersion methods. (A) A two-state exchange between WC and HG A-T and G-C<sup>+</sup> HG base-pairs showing the relative populations and exchange rate constants obtained from  $R_1$  relaxation dispersion experiments at pH 6.8 and the estimated p $K_a$  for a HG G-C+ base-pair. (B) Corresponding thermodynamic profiles obtained from a temperature dependence of  $R_1$  relaxation dispersion at pH 5.4 (G, free energy; H, enthalpy; TS, entropy).