DNA breathes Hoogsteen

Maxim D. Frank-Kamenetskii

Department of Biomedical Engineering; Boston University; Boston, MA USA

A recent claim is discussed that Watson-Crick pairs in the naked duplex DNA spontaneously flip into Hoogsteen pairs under ordinary conditions. The claim is considered within the historical retrospective and is put into the broader context of DNA biophysics.

A Nature paper on spontaneous flipping of normal Watson-Crick (WC) base pairs into Hoogsteen (HG) base pairs in naked DNA under most normal conditions¹ produced a bombshell effect throughout the world community of nucleic acids biophysicists. Just think about it: the paper claims that, on the average, roughly every hundredth base pair in DNA forms an unorthodox HG pair rather than a classical WC pair! How could such a fundamental fact about “the most important molecule” be overlooked for so long? Must not HG pairs in DNA have been discovered or, at least, suggested long ago? And if not, does it not mean that the claim just cannot be true? After all, recent history has taught the scientific community to be very skeptical to any unusual claim, especially appearing in such high-profile journals like *Nature* and *Science*. It suffices to mention the most recent embarrassment with publication in *Science*, with loud fanfares, of an article about “arsenic life”.² Once someone reads carefully the article and supporting data, it becomes clear that the data by no means support the claim that arsenic substitutes phosphorus in DNA of a bacterial strain obtained by the authors.²

Therefore, a priori skepticism with respect of the claim¹ about fluctuational formation to HG pairs is very much justified. But, in sharp contrast with the “arsenic life” paper,² careful reading of

the HG paper convinces that the claim by Nikolova et al. is based on a very solid ground. They used a new technique, based on carbon NMR, while normally proton NMR is used to study DNA structure and dynamics. As positive controls, they studied chemically modified bases, that cannot form WC pairs but can form HG pairs, and showed that they produce NMR signals similar to the signals they assign to HG pairs. Since formation of the HG pair GC⁺ requires protonation, one must expect their formation to be strongly pH-dependent, which was indeed the case.

By themselves, the HG pairs are nothing new in the field of DNA biophysics. They are very well known since early 1960s when they were discovered by Hoogsteen in crystals formed by DNA bases (see Fig. 1A). They play crucial role in triplexes, both inter- and intramolecular, since bases of the third strand of the triplex, which lies in the major groove of the duplex, form Hoogsteen pairs with purines of the duplex (reviewed in ref. 3). They also play an important role in sequence-specific targeting duplex DNA with PNA⁴ and they have been found to be formed in complexes of duplex DNA with some proteins.⁵ However, the spontaneous formation of HG pairs within regular B-DNA under normal conditions would seem highly unlikely due to several reasons:

(1) Since Watson and Crick first suggested complementary base pairs, their isomorphism, i.e., almost identical distance between glycosidic bonds in both pairs (1.1 nm), has always been considered as very important feature of the regular double helix. By contrast, the interglycosidic distance in HG pairs is much smaller than in WC pairs (0.8 nm), which must

Key words: duplex DNA breathing, hoogsteen base pairs, DNA structure, DNA motility

Submitted: 03/14/11

Accepted: 03/17/11

DOI: 10.4161/adna.2.1.15509

Correspondence to: Maxim D. Frank-Kamenetskii;
Email: mfk@bu.edu

Commentary to: Nikolova EN, Kim E, Wise AA, O'Brien PJ, Andricioaei I, Al-Hashimi HM. Transient Hoogsteen base pairs in canonical duplex DNA. *Nature* 2011; 470:498–504; PMID: 21270796; DOI: 10.1038/nature09775

Figure 1. (A) Hoogsteen (HG) AT and GC⁺ base pairs between antiparallel DNA strands. Note that purines in both pairs must assume the unusual syn conformation. (B) Incorporation of HG pairs into the regular double helix. Only HG pairs (light gray) and two adjacent WC pairs (dark gray) are shown. HG pairs are nicely stacked with neighboring WC pairs.

make their incorporation into the B DNA pretty unfavorable.

(2) In triplexes, DNA strands forming HG pairs AT and GC⁺ with each other are parallel. Obviously, this is not possible in case of transient formation of HG pairs in B-DNA, which strands are antiparallel. The formation of Hoogsteen pairs between antiparallel strands is possible only if one of the bases flips into the unfavorable syn nucleoside conformation. The syn conformation is virtually impossible for pyrimidines while possible for purines. It forms in Z DNA, for instance, although it is unfavorable relative to the anti conformation. Therefore, as Nikolova et al. reasonably assume,¹ purines in a fluctuationally formed HG pair must be in the syn conformation. However, this makes the HG pair formation additionally unfavorable.

(3) The HG pair between guanine and cytosine can be formed only if cytosine is protonated in the N3 position (Fig. 1A). Since pK for protonation of free cytosine is 4.2, at neutral pH the formation of the HG pair GC⁺ must be very unfavorable.

It is pretty amazing that in spite of these negative factors, the probability of HG pair formation in DNA is found to be as high as 10⁻², which corresponds to the ΔG value between HG and WC pairs of

less than 3 kcal/mol. Of course, the most important interactions stabilizing the DNA duplex, stacking interactions, can be restored when a WC pair flips into a HG pair, which probably explains its stability (see Fig. 1B).

Thus, the arguments in favor of fluctuational formation of HG pairs in DNA presented by Nikolova et al. look very convincing. Still, the question remains: how can this new finding be incorporated into the body of knowledge about DNA biophysics? Does it contradict to the well established facts about DNA? To answer these questions, let me briefly overview what we knew about DNA fluctuational motility before the *Nature* publication.¹

The fact that the DNA structure cannot be rigid and must experience considerable thermal fluctuations has been appreciated for decades. It has been a mainstream view that the accumulation of small bending fluctuations leads to the observed persistence length of DNA, which behaves as a worm-like chain. Similarly, the accumulation of small torsional fluctuations determines many properties of circular DNA molecules. It has also been suggested that the double helix experiences rarer but much larger fluctuations, dubbed “breathings”. Such breathings or fluctuational openings of the DNA duplex were

invoked to explain chemical modification of the groups in DNA buried inside the duplex and therefore inaccessible when DNA is in the duplex form. Exchange of imino hydrogens in thymine and guanine was the most extensively studied process of this sort. Another phenomenon, which played an important role in elucidating DNA breathing, was the DNA reaction with formaldehyde. Here we do not mean DNA crosslinking by formaldehyde, which is an important but very rare event, but formaldehyde’s chemical modification of amino and imino groups of DNA bases eventually leading to separation of DNA strands or DNA unwinding.

Originally, the data on hydrogen exchange were interpreted as yielding a probability of base pair openings as high as 10⁻². This figure was in a sharp disagreement with the estimate of 10⁻⁵ following from the formaldehyde data. In an obscure but influential publication made from behind the Iron Curtain back in 1985, I emphasized the controversy between the hydrogen exchange and the formaldehyde data and insisted that something was wrong with the interpretation of the hydrogen exchange data although I did not have any idea what exactly was wrong.⁶ The conflict was resolved soon after my publication⁶ by

Gueron et al. who applied proton NMR to study hydrogen exchange in DNA.⁷ The authors convincingly showed that the NMR data on hydrogen exchange yielded an opening probability of 10^{-5} . But what was wrong with the previous hydrogen exchange data? Gueron et al. discovered that in a disrupted base pair one base catalyzes the imino group hydrogen exchange in the other, complementary base. This catalysis highly increases the rate of hydrogen exchange, which leads to a gross overestimation of the DNA opening probability if the catalysis is neglected. In my News and Views comments⁸ to the Gueron et al. paper,⁷ I summarized our knowledge on DNA breathing, circa 1987:

At conditions well below the DNA melting temperature, only separate base pairs open within the double helix with probability of 10^{-5} . They live in the open state for about a microsecond and then close again. Terminal base pairs open with much higher probability.

This picture of DNA breathing has been fully confirmed during the past years in numerous experimental and theoretical studies (reviewed in ref. 9).

Now we have learned that in addition to the deep breath, or full base-pair opening, there happens a smaller but much more probable "Hoogsteen breath". But why has it not been detected by proton NMR? In case of AT pairs the answer is obvious. Indeed, the thymine imino proton participates in hydrogen bonding in both, WC and HG pairs, and therefore is unavailable for exchange in both cases. The situation with GC pairs is much less obvious. The

guanine imino proton does not participate in hydrogen bonding in the HG GC⁺ pair and must be available for exchange. But it is true only for alkaline pH, where HG GC⁺ pairs are not formed anyway because of the requirement for cytosine protonation. At neutral and acidic pH the guanine imino proton exchange is known to be a very complicated process. It goes via protonation at N7 position of guanine and experiences enormous catalysis from the complementary cytosine.¹⁰ The fact that the N7 position is blocked for protonation by HG hydrogen bonding can prevent the normal route of exchange at neutral and acidic pH. So most probably the Hoogsteen breathing could not be detected by proton NMR.

Could Hoogsteen breathing manifest itself in any other ways, rather than in carbon NMR experiments? Now DNA biophysicists have to carefully look at the possible consequences of the new mode of the double helix breathing and check whether the transient formation of HG pairs can result in new effects, which could not be expected before. For instance, if incorporation of a HG pair into B-DNA results in formation of a hinge, significantly more flexible than regular B-DNA, one can try to detect a pH dependence of the DNA persistence length at neutral and acidic pH range. Of course, the effect must be very small and its detection requires extremely accurate measurements of the DNA persistence length, but such measurements are currently possible.

How important is the discovery of Hoogsteen breathing? In my opinion, it is hugely important. The elucidation of the

exact nature of fluctuations in the DNA double helix is crucial for understanding how exactly DNA is damaged by various chemicals, what is the mechanism of DNA binding with various drugs and proteins working on the level of the DNA double helix. There is no doubt in my mind that this exciting development will lead to new discoveries in the field of DNA biophysics and beyond.

Acknowledgement

I thank Dr. Dmitry Beglov for preparing the figure.

References

1. Nikolova EN, Kim E, Wise AA, O'Brien PJ, Andricioaei I, Al-Hashimi HM. Transient Hoogsteen base pairs in canonical duplex DNA. *Nature* 2011; 470:498-504.
2. Wolfe-Simon F, Blum JS, Kulp TR, Gordon GW, Hoelt SE, Pett-Ridge J, et al. A bacterium that can grow by using arsenic instead of phosphorus. *Science* 2010;17(4):423-9. DOI: 10.1126/science.1197258
3. Frank-Kamenetskii MD, Mirkin SM. Triplex DNA structures. *Ann Rev Biochem* 1995; 64:65-95.
4. Demidov VV, Frank-Kamenetskii MD. Two sides of the coin: affinity and specificity of nucleic acid interactions. *TIBS* 2004; 29:62-71.
5. Kitayner M, Rozenberg H, Rohs R, Suad O, Rabinovich D, Honig B et al. Diversity in DNA recognition by p53 revealed by crystal structures with Hoogsteen base pairs. *Nature Struct Mol Biol* 2010; 17:423-9.
6. Frank-Kamenetskii MD. Fluctuational motility of DNA. *Structure & Motion: Membranes, Nucleic Acids & Proteins*. Clementi E et al, eds. Adenine Press, NY 1985; 417-32. For pdf of this article, go to: http://people.bu.edu/mfk/mfk_1985.pdf.
7. Gueron M, Kochoyan M, Leroy JL. A single mode of DNA base-pair opening drives imino proton exchange. *Nature* 1987; 328:89-92.
8. Frank-Kamenetskii MD. How the double helix breathes. *Nature* 1987; 328:17-8.
9. Krueger A, Protozanova E, Frank-Kamenetskii MD. Sequence-dependent base pair opening in DNA double helix. *Biophys J* 2006; 90:3091-9.
10. Nonin S, Leroy JL, Gueron M. Acid-induced exchange of the imino proton in G-C pairs. *Nucl Acids Res* 1996; 24:586-95.