Fluctuations in the DNA double helix: A critical review

Maxim D. Frank-Kamenetskii a,*, Shikha Prakash b

a Department of Biomedical Engineering, Boston University, 44 Cummington Mall, Boston, MA 02215, USA
b Department of Physics, University of Pune, Pune 411007, India

Received 14 October 2013; received in revised form 27 December 2013; accepted 30 December 2013
Available online 13 January 2014
Communicated by E. Shakhnovich

Abstract

A critical overview of the extensive literature on fluctuations in the DNA double helix is presented. Both theory and experiment are comprehensively reviewed and analyzed. Fluctuations, which open up the DNA double helix making bases accessible for hydrogen exchange and chemical modification, are the main focus of the review. Theoretical descriptions of the DNA fluctuations are discussed with special emphasis on most popular among them: the nonlinear-dynamic Peyrard–Bishop–Dauxois (PBD) model and the empirical two-state (or helix–coil) model. The experimental data on the issue are comprehensively overviewed in the historical retrospective with main emphasis on the hydrogen exchange data and formaldehyde kinetics. The theoretical descriptions are critically evaluated from the viewpoint of their applicability to describe DNA in water environment and from the viewpoint of agreement of their predictions with the reliable experimental data. The presented analysis makes it possible to conclude that, while the two-state model is most adequate from theoretical viewpoint and its predictions, based on an empirical parametrization, agree with experimental data very well, the PBD model is inapplicable to DNA in water from theoretical viewpoint on one hand and it makes predictions totally incompatible with reliable experimental data on the other. In particular, it is argued that any oscillation movements of nucleotides, assumed by the PBD model, are severely damped in water, that no “bubbles”, which the PBD model predicts, exist in reality in linear DNA well below the melting range and the lifetime of an open state in DNA is actually 5 orders of magnitude longer than the value predicted by the PBD model.

© 2014 Elsevier B.V. All rights reserved.

Keywords: DNA biophysics; Double helix breathing; Peyrard–Bishop–Dauxois model; Helix-coil model; Hydrogen exchange method; Formaldehyde method

Contents

1. Introduction ................................................................................................................................. 154
2. Theoretical approaches ............................................................................................................. 155
  2.1. Peyrard–Bishop–Dauxois (PBD) model .............................................................................. 155
  2.2. The PBD model predictions ............................................................................................... 157
  2.3. The tyranny of water .......................................................................................................... 158

* Corresponding author.
E-mail address: mfk@bu.edu (M.D. Frank-Kamenetskii).

1571-0645/$ – see front matter © 2014 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.plrev.2014.01.005
1. Introduction

At a first glance DNA looks like an aperiodic one-dimensional crystal where each base pair is flanked only by two neighbors (see e.g.; [1]). But since it is a nanometer-scale object, it is not the static entity that structural pictures suggest, rather it experiences all kinds of thermal fluctuations ('breathes') including local opening of the base pairs. The stability of double-stranded DNA (dsDNA) at physiological temperature is due to the hydrogen bonding between complementary bases and the stacking between neighboring bases. However, these base-stacking interactions are of the order of magnitude of a few $k_B T$ thermal energy [2–4] and the thermal fluctuations can lead (even at physiological temperature) to local and transient unzipping of the double helix [4–7]. In this article we present a comprehensive picture of the DNA fluctuational motility by overviewing both theoretical and experimental approaches developed so far.

Thermal motion plays a crucial role in functioning of biological molecules in general and of DNA in particular [8]. Indeed, in case of the DNA double helix, many reactive chemical groups of bases are buried within the double helix and are inaccessible for external agents, which can attack these groups only when they are exposed to the solution. This is why thermal motion plays a decisive role in a variety of processes of DNA damage by chemical agents which (just like formaldehyde, for instance) react with DNA groups normally buried inside the double helix. Such chemical modifications are known to lead to lethal effects or after some further processes, to mutations [9]. As a chemical agent, formaldehyde has a known toxic, mutagenic and carcinogenic potential. Historically, formaldehyde reaction with DNA, together with the hydrogen exchange, were the first cases for which the role of fluctuational opening of the double helix was well documented. Although these cases were extensively reviewed in the past [7,10,11], we revisit them in the present paper emphasizing recent developments in the field.

It is due to thermal motion, that the DNA double helix experiences smooth bending and torsional deformation. Because of these movements, the double helix forms a polymer coil in solution, may form a supercoil under topological stress, may be wrapped around the histone core in nucleosomes, etc. Very recently, it has been suggested that such smooth bending may become sharp kinks, in the form of fluctuational base pair opening, when DNA is subjected to very strong bending [12].

Quantitative understanding to DNA motility is impossible without an adequate theoretical treatment of fluctuations in DNA. Developing such a treatment is a very challenging task, taking into account that DNA is a pretty complex chemical substance dissolved in water. Over the years, numerous theoretical approaches have been developed based on different basic models of DNA. Unfortunately, in literature competing theoretical treatments sometimes coexist as if in parallel universes: adepts of one approach publish their results over many years never citing or discussing papers published by adepts of other approaches, and vice versa. A major motivation behind the present article was to end this practice. We strongly believe that the field of DNA biophysics will enormously benefit from open and candid discussions of competing approaches. We therefore make in this article significant efforts to analyze and compare various theoretical approaches to study DNA fluctuations.
2. Theoretical approaches

From a theoretical physics viewpoint, DNA is an extremely complicated object. In addition to irregular nature of elementary unit of the DNA chain, the nucleotide, DNA is immersed in water and is surrounded by ions. Therefore, any theoretical treatment of DNA is inevitably based on simplified modeling of the real situation. Over the years, a multitude of such models has been suggested. In this section we will discuss some of these models paying special attention to the central question of adequacy of the models as tools to analyze fluctuations in DNA.

2.1. Peyrard–Bishop–Dauxois (PBD) model

The PBD model, originally suggested by Peyrard and Bishop in 1989 [13], has attracted a lot of followers (reviewed in [14]). Within the PBD model, nucleotides are treated as beads connected with two types of nonlinear springs, corresponding to interactions between bases within the base pair and stacking interaction between adjacent base pairs along the DNA chain (see Fig. 1). The water around DNA is totally ignored so the model considers DNA as if it were in vacuum. It is very easy to write the dynamic equations for such a model. The main characteristics of PBD model are:

- The longitudinal displacements of bases are neglected as their amplitude of vibrations is much smaller than the amplitude of vibration of transverse motion. The displacements from the equilibrium of the \( n \)th nucleotide are denoted by \( u_n \) for one chain and \( v_n \) for the other.
- Two neighboring nucleotides of the same strand are connected by a harmonic potential (later on this term was modified by Dauxois to an anharmonic potential [15]).
- The hydrogen bonding between two bases on opposite strands is represented by the Morse potential. It represents not only the attraction between base pairs but also the repulsion between the two phosphate groups on the opposite strands.

The Hamiltonian [16] of the system can be written as,
\[ H = \sum_n \left[ \frac{1}{2} m \left( \dot{x}_n^2 + \dot{y}_n^2 \right) + \frac{1}{2} k \left( x_n - x_{n+1} \right)^2 + \frac{1}{2} k \left( y_n - y_{n+1} \right)^2 + D \left( e^{-\alpha (y_n - y_{n+1})/\sqrt{2}} - 1 \right)^2 \right] \]  

(1)

Here \( m \) is the reduced mass of the base pair and \( k \) is the force constant. The first term in the above equation represents the kinetic energy of the system and the second term is the stacking potential which represents the effect of stretching of a pair on nearest neighbor positions. The third term is the Morse Potential which represents the interaction between two bases. In reduced coordinates the motions of the two strands can be described as,

\[ x_n = \frac{u_n + v_n}{\sqrt{2}}, \quad y_n = \frac{u_n - v_n}{\sqrt{2}} \]

which represents the in phase and out of phase motions, respectively. In terms of reduced coordinates, the Hamiltonian can be written as,

\[ H = \sum_n \left[ \frac{1}{2} m \dot{x}_n^2 + \frac{k}{2} (x_n - x_{n+1})^2 \right] + \sum_n \left[ \frac{1}{2} m \dot{y}_n^2 + \frac{k}{2} (y_n - y_{n+1})^2 + D \left( e^{-\alpha y_n} - 1 \right)^2 \right] \]

(2)

The part of the Hamiltonian which depends on the variable \( x_n \) is decoupled from stretching part and corresponds merely to a harmonic chain without on-site potential. One can ignore this term in studying the opening of the DNA helix. The effective Hamiltonian will be:

\[ H = \sum_n \left[ \frac{p_n^2}{2m} + W(y_n, y_{n+1}) + V(y_n) \right] \]

(3)

where:

\[ W(y_n, y_{n+1}) = \frac{k}{2} (y_n - y_{n+1})^2 \quad \text{and} \quad V(y_n) = D \left( e^{-\alpha y_n} - 1 \right)^2 \]

A harmonic stacking potential like in Eq. (3) results in a very wide melting range for DNA, which disagrees with the experiment. Based on this argument, Dauxois et al. [15] modified this potential and a nonlinear term was introduced. This nonlinear term takes care of the cooperativity of melting and changed the nature of transition from smooth to very sharp. Since stacking energy is not the property of individual bases but is due to base pairing, when the hydrogen bonds between the bases stretched or break, the electronic configuration of the bases modifies which reduces the coupling between the neighboring bases. The modified stacking potential can be written as,

\[ W(y_n, y_{n+1}) = \frac{k}{2} \left( 1 + \rho e^{-\alpha (y_n + y_{n+1})} \right) (y_n - y_{n+1})^2 \]

(4)

The nonlinear coefficient to \((y_n - y_{n+1})^2\) depends on the sum of the stretchiness of the two interacting base pairs and decreases from \(\frac{1}{2}k(1 + \rho)\) to \(k/2\) when either one (or both) base pair is stretched. Dauxois et al. [15] also provided with the cooperative effects that were introduced phenomenologically, within the two-state model (see Section 2.6). One can visualize that a base pair in the vicinity of an open site has lower vibrational frequency which reduces its contribution to the free energy but lower coupling along the strands gives bases higher degree of freedom to move independently from each other which increases the entropy of the system.

The on-site potential not only represents the coupling between two bases on opposite strands but also the repulsive interaction between the phosphate groups of two strands. The on-site potential (given in Eq. (3)) is represented by the Morse potential [17], where \(D\) is the depth or height of the potential, \(y\) is the displacement between atoms, \(a\) is an inverse width of potential well. Any other potentials, like Lennard-Jones potential, can be used to represent the on-site potential in DNA, but Morse potential is found to be a good representation of the energy as a function of end atom distance when the bonds are hydrogen bonds [18].

Therefore, the complete Hamiltonian is:

\[ H = \sum_n \left[ \frac{p_n^2}{2m} + \frac{k}{2} \left( 1 + \rho e^{-\alpha (y_n + y_{n+1})} \right) (y_n - y_{n+1})^2 + D \left( e^{-\alpha y_n} - 1 \right)^2 \right] \]

(5)

The PBD model described above can be viewed as a model of a one-dimensional monoatomic lattice (see Fig. 1) with each atom having mass \(m\) and nearest neighbor interaction given by \(W(y_n, y_{n+1})\) (i.e. the stacking interaction). Each “atom” of the lattice is subjected to an “external” potential given by \(V(y_n)\) (i.e. Morse potential) which effect
is to confine the chain in the potential well. The melting transition takes place because of competition between the thermal energy which leads to displacement of atoms from potential minimum and the combined effect of “external field” and the nearest neighbor interactions which lead to the confinement. The model is different from the other usual one-dimensional models that do not show phase transition.

Though, as described above, the PBD model ignores the helical structure of dsDNA, it is found to include enough details to analyze the behavior at a few Å scale relevant to molecular-biological events.

2.2. The PBD model predictions

Let us now consider how fluctuational openings of base pairs are calculated within the PBD model. In the model, nth base pair is considered open when its stretching coordinate, \( y_n \), exceeds the threshold value, \( y_0 \). For a sequence of \( N \) base pairs with periodic boundary conditions, partition function can be written in terms of Hamiltonian given in Eq. (5) as:

\[
Z = \int \prod_{n=1}^{N} dy_n dp_n \exp[-\beta H(p_1 - p_N)\delta(y_1 - y_N)] = Z_p Z_y
\]  

(6)

The momentum part is readily integrated to give the familiar kinetic-energy factor for \( N \) particles \( Z_p = (2\pi m k_B T)^N/2 \). Since the coupling involves only nearest-neighbor interactions, \( Z_y \) can be expressed in the form

\[
Z_y = \int \prod_{n=1}^{N} dy_n \exp[-\beta H(y_n, y_{n+1})]\delta(y_1 - y_N)
\]  

(7)

The quantity which gives average thermal stretching of a base pair, is calculated by:

\[
\langle y_n \rangle = \frac{1}{Z_y} \int \prod_{n=1}^{N} dy_n e^{-\beta H(y_n, y_{n+1})}\delta(y_1 - y_N) dy_n
\]  

(8)

As the model is assumed to be homogeneous, the result does not depend on the particular site \( n \) considered. The integral can again be calculated with the Transfer Integral (TI) method which gives [19]:

\[
\langle y \rangle = \langle y_n \rangle = \int \phi_0^2(y) y dy
\]  

(9)

\( \langle y \rangle \) is used to calculate the fraction of open base pairs. The nth base pair is considered to be in open state if \( \langle y \rangle \) exceeds a certain threshold. Initially proposed values of Morse potential parameters are \( D = 0.33 \) eV and \( a = 1.8 \) Å\(^{-1}\), which corresponds to mean values for the N–H⋯N and N–H⋯O bonds in the A–T and G–C base pairs. On calculation this shows a rapid increase of \( \langle y \rangle \) around a particular temperature which is a characteristic of DNA melting. The results are very sensitive to the parameters of the hydrogen bonds which bind the two strands and to the intrastand interaction constant. The increase in \( k \) increases the melting temperature and it is suggested that \( k \) must be of the order of \( 3.0 \times 10^{-3} \) eV/Å\(^2\) to obtain a reasonable melting temperature. The stacking interaction in the first attempts was purely harmonic (\( \rho = 0 \)) [13], but it decreases when the complementary bases reach farther (positive): this nonlinear term was found to be relevant to give cooperativity to the melting process. To model the heterogeneous DNA’s two different values of \( D_i \) have been inserted. Since A–T has two hydrogen bonds, while the G–C has three, the depth for the G–C Morse potential is 1.5 times of A–T. By adjusting the model parameters with respect to experimental results, the complete set of parameters are \( D_{GC} = 0.075 \) eV and \( a_{GC} = 6.9 \) Å\(^{-1}\) for the G–C base pair and \( D_{AT} = 0.05 \) eV and \( a_{AT} = 4.2 \) Å\(^{-1}\) for the A–T base pair and for the inter site coupling \( \rho = 2.0, \alpha = 0.35 \) Å\(^{-1}\) and \( \kappa = 0.025 \) eV/Å\(^2\) [20].

The lifetime of bubbles are studied in Ref. [21], where the thermal dynamics of the nth base pair is obtained through

\[
m \ddot{y}_n = -V'(y_n) - W'(y_{n+1}, y_n) - W'(y_n, y_{n-1}) - m \gamma \dot{y}_n + \xi_n(t)
\]  

(10)

Here a prime denotes differentiation with respect to \( y_n \), \( \gamma \) is a friction constant whose value is 0.05 ps\(^{-1}\) (\( \gamma = 0.005 \) ps\(^{-1}\)) during (after) preheating.
Initially, the stacking interaction was considered sequence insensitive and sequence only affected the on-site Morse potential. A more accurate description of the stacking interaction requires consideration of all possible combinations of neighboring base-pairs. Sequence-dependent stacking term has been included in the original PBD model by Alexan-
drov et al. in 2009 [22].

2.3. The tyranny of water

The PBD model is extremely attractive to physicists since it requires little knowledge about DNA biophysics to predict, essentially from first principles, the behavior of DNA with respect to base pair opening and closing. Not surprisingly, Peyrard and Bishop’s 1989 paper, which first introduced the model [13], was hailed in an extended commentary by the legendary Nature editor at that time, Sir John Maddox [23]. Although the original model did not take into account the DNA sequence considering all elementary units of the DNA chain as identical, the model was later revised to include such differences [22]. There are a series of publications where the predictions based on the PBD model are correlated with biological experiments (see [24] and references therein). Moreover, very recently one of fundamental predictions of the model, movement of DNA nucleotides in the ps range, has been claimed to be experimentally confirmed by the effects of the THz radiation on stem cells [25].

However, in our opinion, before applying the model to analyze biological in vitro experiments, let alone living cells, one has to check how well the predictions of the model correlate with biophysical experimental data on DNA. We will address this question below in Section 4.1. In this section, we raise an even more fundamental issue, which, to the best of our knowledge, has not been addressed in literature. It is that we believe that the very applicability of the PBD model to DNA, from the theoretical viewpoint, is very much debatable because the model totally ignores water.

Introduction of water fundamentally affects the original model in two ways. First, nucleotide vibrations predicted by the model must be severely damped in the water medium. Secondly, in open bases the groups capable of forming H-bonds must attract water molecules and these water molecules have to dissociate before complementary bases could reform base pairs. Both effects must profoundly affect the picture of nucleotide anharmonic vibrations described by a Hamiltonian like in Eq. (5) The effect of hydration of bases in the open state is pretty obvious and it leads to the activation barrier for base pairs opening and closing. There were attempts in the literature to incorporate such a barrier into the PBD model [26–31]. Below, we concentrate on the damping effect of water because of its paramount importance and because it has never been addressed before in the vast literature on the PBD model.

It is well known that at nanoscale in water there is no inertia: the in vacuo Newtonian dynamics approach described by equations like Eq. (5) is not valid (see, e.g. [8,32]). Indeed let us assume that, in the relevant dynamic scale, water can be treated as a continuous liquid medium. Let us consider a simple problem: dynamics of a sphere of radius $r$ and mass $m$ in water moving with velocity $v$. The Stokes law leads to the following equation:

$$m\dot{v} = -6\pi r v \eta$$

(11)

where $\eta$ is the dynamic viscosity of water. Eq. (11) yields an exponential decay:

$$v = v_0 \exp(-t/\tau)$$

(12)

with

$$\tau = \frac{2}{9} r^2 (\rho_b/\eta)$$

(13)

where $\rho_b$ is the density of the bead material.

If we assume $r = 1$ nm, we will get $\tau = 0.5$ ps. If we assume that the bead experiences thermal motion at ambient temperature, we obtain that the bead will shift at the distance of 0.02 nm (1/100 of its diameter!) for 0.5 ps, the characteristic damping time [8,32]. In other words, vibrations in water at nanoscale are definitely not like macroscopic pendulum’s oscillations in vacuum, and even not like macroscopic pendulum’s oscillations in water (which are already damped pretty fast) but rather they are like macroscopic pendulum’s oscillations in honey (see [32], p. 8).

The above consideration assumes very slow movement of the sphere when the Stokes law is valid. Let us consider the opposite limiting case of very high frequency vibrations of the sphere [33].

In the high-frequency limit the equation of motion of an oscillating sphere is [34]:
\[ \dot{u} = -\frac{3\pi r^2 \sqrt{2\eta\rho_0}\omega}{M}u \]  

(14)

where \( M = m + m_f / 2 \) and \( m \) is mass of the sphere and \( m_f \) is mass of the displaced water, \( \rho_0 \) is density of water and \( \omega \) is the angular frequency of the bead’s vibration.

The relaxation time is:

\[ \tau = \frac{M}{3\pi r^2 \rho_0 \sqrt{2\nu\omega}} = \frac{4M}{9m_f \sqrt{2\nu\omega}} \]  

(15)

where \( \nu \) is kinematic viscosity of water, \( \nu = 10^{-2} \text{ cm}^2 \text{s}^{-1} \).

If we take \( m_f \approx M, r \approx 10^{-7} \text{ cm} \) and \( \omega \approx 10^{12} \text{ s}^{-1} \) (the typical figure for the bead’s vibrational frequency): we will get \( \tau \approx 10^{-12} \text{ s} \).

We see that despite very different physics, both limiting cases (very slow and very fast movement) yield the same estimate for the relaxation time of a nanoscale bead in water: in the range of ps. As the result, no vibrations are possible: the particle’s kinetic energy fully dissipates while the particle moves at the distance, which is two orders of magnitude shorter that the particle’s diameter (see above). We conclude that fluid mechanics predicts the complete damping of any vibrations of nanoscale objects in water. The significance of this conclusion for the entire field of molecular biophysics cannot be overestimated. It has truly fundamental consequences propagating far beyond the scope of the current article. Among many others, it dismisses inertia as a player in functioning of molecular motors \[8,32\] and it dismisses a popular notion of “Davydov solitons” as the energy transfer mechanism in proteins \[35\]. And, of course, it dismisses the PBD model for the analysis of DNA fluctuations.

One can argue that the above reasoning is based on a continuous medium approach, which is a crude approximation for the nanoscale where molecular structure of water must be allowed for. It is true, of course, but the above reasoning clearly indicates that the vacuum approximation is definitely much worse than the continuum medium approximation. In any dynamic model, the tyranny of water cannot be avoided just by claiming that water is allowed for in parameters of the Hamiltonian of the type of Eq. (5). Therefore, water cannot be neglected in any dynamic model of DNA: it must be allowed for either as a continuous medium or water must be included explicitly into the dynamic equations, as it is done in the molecular dynamics (MD) simulations (see below).

2.4. Molecular dynamics

Since, as we argued in the previous section, water cannot be neglected, it is natural to turn to a method, which treats DNA and surrounding environment in their full complexity. Such brute-force approach to theoretical modeling of biomolecular structure and dynamic is known as Molecular Dynamics (MD). In application to DNA, it means to consider DNA in its full complexity to include explicitly surrounding water molecules and ions. Then the full set of classical equations of motion with potentials between atoms are considered and solved using powerful computers. Obviously, it is an extremely challenging task: not only huge computational problems must be overcome but interaction potentials between atoms must be determined with a pretty high accuracy. For recent review of the area of DNA MD simulations, see Ref. \[36\]. Any specific realization of this program, however, still raises many questions. In MD simulations, classical equations of motion are considered and empirical potentials are included. The fully consistent program, from theoretical viewpoint, would consist in solving the Schrödinger equation for the whole system consisting of DNA, water and ions. Quantum mechanical approaches are used for separate DNA components, like bases or even base pairs, but they are still unfeasible for the full DNA molecule, even very short one \[37\].

Still, MD and related methods gain more and more popularity and some attempts to consider the pathways of base pairs opening in DNA have been undertaken (see \[38\]). However, even if the MD method yielded very reliable results, which could be fully trusted (which is not the case), we cannot expect to observe any fluctuational opening events during MD simulations. Indeed, because of extreme computational complexity of MD simulations, they normally can be performed for duration in the ns range, more recently up to \( \mu \text{s} \) \[36\]. But as experimental data indicate (see Section 3), the lifetime of closed base pair state is much larger: in the range of 10 ms. We therefore conclude that one cannot expect any reliable data on base-pair fluctuational opening from MD simulations.
2.5. Langevin and Brownian dynamics

In a simplified version of MD called Langevin dynamics, water is treated as the continuous medium with the correct macroscopic parameters but still dynamic equations are considered with all participants experiencing random hits simulating the thermal motion [39]. More logical in case of DNA is to use even more simplified version, called Brownian dynamics, where the acceleration term is totally eliminated from the Langevin equations of motion.

Although Brownian dynamics is an appropriate tool to simulate movements of DNA as a macromolecule (see, e.g., [40,41] and references therein), this approach cannot be consistently applied to base pair opening process since in the open state interaction of water molecules with bases plays a crucial role.

2.6. The two-state model

As we have seen above, all dynamic approaches to treat breathings of the DNA double helix encounter serious difficulties. The brute-force full-blown MD approach still has little chances to succeed in quantitative predictions of DNA fluctuations because of enormous computational difficulties and because we do not know (and may never know) potentials with sufficient precision. All simplified models (PBD, Langevin dynamics, Brownian dynamics, etc.) severely suffer from the water tyranny: on one hand you cannot neglect water and on the other hand you cannot treat it as a continuous medium because water structure is very important at the sub-nanoscale; and when you start modify your simplified models making them more realistic, you asymptotically approach the MD, with all its problems.

Therefore, it goes with no surprise that the most successful theoretical approach to treat DNA fluctuations still remains an empirical one, based on the most simple two-state model. The two-state model emerged from the theoretical treatment of DNA melting, or helix–coil transition [7,42–44], and it is often called also “helix–coil” model [45] or “Poland-Scheraga” model [46].

Within the framework of the model, each base pair is assumed to be in one of two states: closed or open (Fig. 2). In its simplest version, a homopolymer without loop factors, the state of DNA consisting of the sequence of open and closed base pairs \( k_1, k_2, \ldots, k_N \) (where \( k = 1 \) for open state and \( k = 2 \) for closed state) is described by the following Hamiltonian [7,42–44]:

\[
H = \sum_{n=1}^{N} (F \delta_{k_n2} - F_x \delta_{k_n2} \delta_{k_{n+1}1})
\]

where \( \delta \) is the Kronecker delta, \( F \) and \( F_x \) are empirical parameters of the model, which are supposed to be determined from comparison of theory with experiment. Note that the Hamiltonian in Eq. (16) is fully equivalent to the
Hamiltonian for the one-dimensional Ising model of spins, with spin up corresponding to open state and spin down corresponding to closed state [47].

The two-state model is obviously immune to the tyranny of water since the empirical parameters automatically allow for water and small ions effects when they are determined via comparison of the theory with experiment performed under given conditions. Still, the model in its simplest form as in Eq. (16) cannot adequately describe the real DNA. Two major effects must be taken care of in the comprehensive theory of DNA melting: dependence of Hamiltonian on the nucleotide sequence and the entropy of loops formed by melted regions [48–50].

An enormous advantage of the two-state model consists in the fact that over the years the DNA melting theory based on the model has been perfected and extensively used to analyze the data on DNA melting so that it quantitatively explains DNA melting in great details [2,48–50]. The theory starts from a Hamiltonian of a type as in Eq. (16), which also takes into account the DNA sequence and the fact that melting regions in DNA form closed loops. The Hamiltonian is used to calculate the partition function of DNA, which, for Hamiltonian in the form of Eq. (16), is [7,42–44]:

$$Z = \sum_{k_1=1,2} \cdots \sum_{k_N=1,2} \exp \left\{-\left(\frac{1}{k_BT}\right) \sum_{n=1}^{N} (F\delta_{k_n^2} - F_s\delta_{k_n^2}\delta_{k_{n+1}})\right\}$$

(17)

From the partition function the melting curve (the fraction of open base pairs as the function of temperature) and other thermodynamic parameters of DNA are calculated using rigorous computer algorithms developed in the field. Very early, it was suggested to theoretically predict the opening probability by directly applying the DNA melting theory to temperatures far below the DNA melting range [7]. Such extrapolation yielded a clear prediction that only individual base pairs must be open well below melting range and the opening probability must be about $10^{-5}$ [7,9,51]. However, in 1970s, experimental studies, based on the DNA hydrogen exchange and formaldehyde reaction with DNA, mostly testified in favor for much larger base-pair opening probability: $10^{-2}$ (see Section 4). This casted serious doubts on the validity of the DNA melting theory at temperatures outside the melting range, where it was fully validated, the opinion, which became a conventional wisdom in the field and even found its way to, now classical, Cantor and Schimmel’s textbook [52] pp. 1170–3. A minority opinion, consisting in assertion that the correctly analyzed formaldehyde data were actually in agreement with the two-state model prediction [9,10,53] and that something was wrong with H-exchange data analysis [10], was mostly ignored. The two-state model was fully exonerated as a predictor the DNA base pair opening probability only after a seminal NMR study by Guéron et al. [54], as will be discussed at length below (Section 3.3).

In contrast to dynamic theories discussed in previous sections, the DNA melting theory is purely thermodynamic and cannot predict the lifetime of open state. However, since classical paper by Glauber on the Ising model relaxation [55], it has been well known that the two-state model is very much amendable for consideration of relaxational processes. In application to DNA, it was successfully done [56–58]. Corresponding parameters of a simple kinetic model, describing base pair opening and closing, were determined from comparison of theory with experiments for melting of short DNA helices performed using temperature-jump and stop-flow techniques [56,57]. As a result, it was determined that the lifetime of the individual base pair in the open state is in the microsecond range. Respectively, the lifetime of any base pair in the closed state appears to be in the range of $10^{-1}$ to $10^{-2}$ s [10].

3. Experimental approaches

The study of the phenomenon of life is an inherently experimental endeavor. It is because in biology we study not general laws of nature, like in physics, but rather a specific form of life at one separate planet in the entire Universe. And since life, as we know it, is the result of 3.5 billions of years of evolution, which is a stochastic process of trial and errors, only experiment can say what are the elements of life today and how they are constructed. Life cannot be deduced from the first principles and there are no universal laws in biology [59].

Therefore, although various simplified models of biologically significant molecules, like DNA, can present some interest from the pure physical viewpoint, only those of them, whose conclusions are validated by adequate experimental data, are of any interest from biological perspective. In this section, we consider main experimental approaches, which have been used to study DNA fluctuations, and formulate the conclusions derived from these studies. Then we will be in position to discuss various theoretical models we considered above in the light of the experimental evidences.
3.1. Hydrogen exchange

Traditionally, hydrogen exchange (HX) has been a method of choice to study structural fluctuations in proteins and nucleic acids (reviewed in [52,60]). In early studies, kinetics of exchange of proton with deuteron in DNA was observed by detecting, in stop-flow experiments, a small shift of the DNA UV absorption spectrum due to the exchange (see [60] and references therein). However, the method did not allow to resolve very fast exchange times. Since mid-1980s, Nuclear Magnetic Resonance (NMR) has been widely used to study fast proton exchange rates (see [11]). In contrast to the stop-flow method, which studies HX in long DNA molecules, the NMR method is applicable only to short DNA oligomers where signals from separate protons can be resolved, assigned and quantitatively studied.

The rational behind the study of DNA fluctuations by the HX method was very simple: If the proton in question cannot be exchanged within the intact double helix, the experimentally observed exchange must be only due to DNA fluctuations (or RNA fluctuations, in case of the RNA duplex). The most obvious candidates for such structure-dependent protons were imino protons of thymine and guanine (Fig. 3). It was very hard to imagine how these protons could exchange with water protons (or deuterons) without base pair opening. Therefore, there was no controversy that the overall HX reaction for imino protons in case of, say, AT base pairs had to look like [11,52,60]:

\[
\text{TH} \cdots \text{A} \overset{(i)}{\rightarrow} \text{TH} + \text{A} \overset{(ii)}{\rightarrow} \text{TH}^* + \text{A}
\]

where H* is either another proton or deuteron. The controversy was about the rate-limiting step of the overall reaction in Eq. (18): whether it was the opening event (i) or the HX event (ii). But before discussing this controversy, which played a crucial role in the determination of the parameters of the base pair opening, let us turn to the other approach to study DNA fluctuations, which appeared on the scene in parallel with the HX method.

3.2. Formaldehyde kinetics

It was discovered pretty early after the discovery of the DNA double helix that in the presence of formaldehyde DNA gradually unwinds yielding the two separate strands. Studying DNA reaction with formaldehyde, it was determined that formaldehyde hydroxymethylates imino and amino groups of bases according to the following reactions:

\[
\text{N} - \text{H} + \text{CH}_2\text{O} \leftrightarrow \text{N} - \text{CH}_2\text{OH}
\]

Again, like in case of HX, it was assumed from the very beginning that DNA unwinding by formaldehyde cannot proceed without some sorts of fluctuations in DNA double helix (reviewed in Refs. [7,10,51,53]). It should be noted...
that we are talking about reversible reaction of formaldehyde with imino and amino groups, not about a very rare event of formation of crosslinks in DNA by formaldehyde.

Since, in contrast to HX, the reaction with formaldehyde leads to new chemical species, hydroxymethylated bases, which cannot reform the same base pairs as unmodified bases, it became clear very early that the comprehensive analysis of the data on DNA unwinding by formaldehyde would not be possible without detailed theoretical modeling of the process. Such modeling became possible after a comprehensive study of kinetics of formaldehyde reaction with nucleotides was performed by McGhee and von Hippel [61,62], and was first performed by Lukashin et al. [9]. It was based on the two-state DNA model with parameters determined from DNA melting studies. Subsequently, more experimental and theoretical studies led to more detailed modeling, which included the melting parameters for hydroxymethylated base pairs and allowed for the “outside” reaction of amino groups of bases within the double helix (see [10] and below).

3.3. Controversy around the DNA base pair opening parameters

In an early analysis of the HX data, the base-pair opening was considered as the rate limiting step in Eq. (18) for imino groups. This conviction was based on the fact that the stop-flow data showed no dependence of the HX rate on the HX catalyst concentration in case of imino protons [60,63]. Indeed, if the opening were not the rate-limiting step, the overall HX rate would depend linearly on the catalyst concentration when the concentration is low since the reaction in Eq. (18) yields the following equation for the exchange rate constant, \( k_{ex} \) [63]:

\[
\frac{1}{k_{ex}} = (Kk_{ch}C)^{-1} + \frac{1}{k_{op}}
\]

where \( K \) is the equilibrium constant of base-pair opening, \( k_{ch}C \) is the HX rate constant for a free nucleotide at concentration of catalyst \( C \), \( k_{op} \) is the rate constant of base opening. The data for different duplexes (DNA and RNA) showed that the \( k_{ex} \) value for imino protons was independent of \( C \) [52,60,63]. Therefore, the conclusion was made that the experimentally observed \( k_{ex} \) value for imino protons was equal to \( k_{op} \).

By contrast, the HX rates for amino protons perfectly obeyed Eq. (21) [60,63]. Therefore, assuming that the same open state was required for exchange imino and amino protons (not an obvious assumption, see below), the full set of parameters for base pair fluctuational opening could be obtained: the \( k_{op} \) value directly from HX rate for imino protons; the equilibrium constant value, \( K \), from HX rates for amino protons as function of the catalyst concentration (using Eq. (21) and HX rates for free nucleotides, \( k_{ch} \)); the rate constant of base-pair closing, \( k_{cl} \), from the relation: \( k_{cl} = k_{op}/K \).

The resulting set, however, was puzzling: \( k_{op} \approx 1 \text{ s}^{-1}, K \approx 10^{-2}; k_{cl} \approx 10^2 \text{ s}^{-1} \) [52,60,63]. Both, very high value of \( K \) (3 orders of magnitude higher than predicted by the helix–coil model, see Section 2.6 above) and very long lifetime of open state (4 orders of magnitude higher than the value expected from kinetic experiments with oligonucleotides, see Section 2.6 above), were very confusing. As a result, a very provocative suggestion was made by Englander et al. that the open state in DNA is actually a soliton moving along DNA as the tsunami wave moves in the ocean [64]. However, like the later proposed PBD model, which was inspired by the Englander et al. soliton hypothesis, the theoretical model of Englander et al. fully ignored water. Using the same analogy we used above in Section 2.3, we can say that solitons in DNA are as possible as tsunami wave in the ocean filled with honey instead of water. To their credit, Englander and Kallenbach quickly retracted their claim about solitons in their seminal review article on hydrogen exchange by stating on p. 589 on the soliton model [60]: “This kind of model now seems unlikely to be involved in HX from nucleic acid duplexes, since oligomeric duplexes exhibit exchange rates similar to polynucleotides...”. Still, the puzzling parameter values for the open state of base pairs in DNA remained for a while in the mainstream of DNA biophysics and found their way into the fundamental text [52].

Meanwhile, on the formaldehyde front the situation was no less confusing. Based on the same data, two groups arrived at conflicting conclusions about the base pair opening probability. McGhee and von Hippel [65] claimed that the formaldehyde data yielded the equilibrium constant of base-pair opening \( K \) in the range of \( 10^{-2} \), the value very close to one obtained later from the HX data (see above). Grounding their theoretical modeling on the two-state model, Frank-Kamenetskii and co-workers insisted that the formaldehyde data could be quantitatively explained on the basis of the model, which predicted the \( K \) value in the range of \( 10^{-5} \) [9,10,51,53]. Frank-Kamenetskii sharply criticized the McGhee and von Hippel paper [65] for neglecting the reaction with thymine imino group while quantitatively
analyzing the data for the DNA duplex [9,53]. He argued, on the basis of theoretical modeling performed in his group, that, although the thymine imino group reaction had very low equilibrium constant, it was important for the overall process and it had been this reaction’s neglect that led McGhee and von Hippel to a highly overestimated value of $K$ [66]. At the same time period, Chay [68] performed a comprehensive theoretical analysis of the data by Utiyama and Doty [69] on unwinding by formaldehyde of synthetic DNA consisting of only A–T pairs at alkaline pH, where reaction with thymine imino groups dominates the process. She arrived at the $K$ value of $10^{-5}$, fully consistent with the figure claimed by Frank-Kamenetskii’s group.

Thus, by mid-1980s the opinion about a very high probability of base pair opening, in the range of $10^{-2}$, dominated the field although dissenting voices were also heard. In addition to arguments based on the formaldehyde data and the two-state model [9,10,51,53,68], researchers involved in studies of the DNA behavior as a macromolecule were very unhappy with such high probability of “hinges” in DNA which had to manifest themselves in changing the DNA persistence length in a way, which contradicted experimental observations [70,71]. However, while the exotic explanations, like solitons, were quickly abandoned by the biophysical community, the high probability and high lifetime of the open state followed from the HX data remained unexplained (see discussion of the issue in Ref. [10]).

3.4. A comprehensive picture of DNA fluctuational opening

The controversy in the field of DNA fluctuational opening was finally resolved by Guéron et al. in their 1987 Nature paper [54]. They overthrew the cornerstone of all previous analyses of the HX data, that base pair opening was the rate limiting step for the exchange of imino protons. They studied short DNA helices, which allowed them to accurately determine the lifetime, in a very wide range of times, of individual base pair openings by measuring imino protons NMR line width. They discovered that the imino proton HX rates actually depended on the catalyst concentration but at higher concentrations than had been used by the Englander’s group [60,63]. Guéron et al. presented convincing evidence that the lack of action of low concentrations of added catalysts is due to the internal catalyst: this role was played in the open state by a complementary base[54].

They obtained the accurate data on the dependence of imino proton exchange rate on the catalyst concentration and, using Eq. (21), determined the equilibrium constant of base-pair opening, which proved to be $10^{-5}$, in full agreement with the prediction of the two-state model and the formaldehyde data (in the Frank-Kamenetskii’s group version) [54, 72]. Extrapolating to very high catalyst concentrations, Guéron et al. determined, using again Eq. (21), the $k_{op}$ value, which proved to be in a range of $10^2$ s$^{-1}$, also in good agreement with the two-state model prediction (see Ref. [72] and Section 2.6).

Therefore, since 1987 the paradigm in the field of DNA biophysics has been that individual base pairs are fluctuationally open well below the DNA melting range and that the opening probability is in the range of $10^{-5}$, that each base pair opens once per 10 ms and the base pair remains in open state for about $10^{-7}$ s [72]. All controversies of the previous period were resolved and the claims, which had puzzled the biophysical community in previous years, proved to be the result of the misinterpretation of the data. Since then, a huge body of data on DNA fluctuational openings have been accumulated (see [4] and references therein). However, the question, which remains unanswered, is the structure of the open state. We still do not know whether open bases form stacking with adjacent base pairs or are totally unstructured [4]. Most probably, the open state includes a variety of microstates but this issue still awaits its comprehensive study.

It must be emphasized that we concentrate in this review on the parameters of DNA fluctuations, which occur well below the melting range, which definitely correspond, under physiological ionic conditions, to room or ambient temperatures. At elevated temperatures close to melting range, in so-called premelting zone, the character of fluctuations significantly changes. This issue was reviewed by Palecek [73] and discussed at some length by Lukashin et al. [9]. We do not consider the premelting phenomena in this article.

4. Different models of DNA fluctuations in the light of experimental evidences

The overview of the experimental situation given in Section 3 shows that there was a period of turmoil in the area prior to 1987 and that for a number of years all kinds of claims circulated in the literature, which did not survive any further in the mainstream of DNA biophysics. Astonishingly, however, these claims have been circulating until today in numerous publications (it is suffice mentioning that, according to the Web of Science, the 1980 soliton paper [64],
which central claim was retracted by the authors in 1983 (see [60] and Section 3.3) has been cited 300 times since 1984). Our major motivation in writing this article was to sort out what is true from what is untrue in the field of fluctuations in DNA.

As we have seen in Section 2, theoretical approaches to predict DNA fluctuations span from the full-scale MD simulations of DNA in all its complexity with surrounding water molecules and ions explicitly allowed for, to all kinds of simplified models among which the PBD model (Sections 2.1, 2.2) and the two-state model (Section 2.6) have gained most popularity. Historically, the two-state model was the first to be applied since it had been remarkably successful in describing DNA melting (see Section 2.6). As have been narrated in Section 3, it was the two-state model, which helped to clean the mess in the field of DNA fluctuations in mid-1980s, from the theoretical perspective. It goes with no surprise, therefore, that the basic features of base pairs opening obtained experimentally happened to be in full agreement with predictions of the two-state model. A very good agreement between theory, based on the two-state model, and the HX and formaldehyde experimental data was reached by 1987 [72]. Most recent analysis of the HX NMR data based on the two-state model has been done by Krueger et al. [4] and the formaldehyde data has been recently revisited by Bohnuud et al. [74], see also Section 4.2 below.

4.1. The relation of the PBD model to experiment

Let us now discuss the PBD model in the light of experimental situation outlined in Section 3. The primary motivation for Peyrard and Bishop to formulate a nonlinear dynamic model of DNA [13], known now as the PBD model, was the soliton hypothesis by Englander et al. [64]. In their 1989 paper [13], however, Peyrard and Bishop failed to mention that in the 1983 review article Englander and Kallenbach retracted the soliton explanation for the anomalously long lifetime of the open state in DNA (in the range of $10^{-2}$ s) because this claim had not survived the experimental test (see [60] and Section 3.3 above). Moreover, as we narrate in Section 3.4, two years before the Peyrard and Bishop’s publication, Guéron et al. in their Nature paper resolved the mystery of long lifetimes of the open state showing that it had been the result of misinterpretation of the HX data by Englander and co-workers [54] and that the lifetime of open state is actually in the range of $10^{-7}$ s rather than $10^{-2}$ s, in full agreement with the predictions based on the two-state model. Astonishingly, however, even as recently as in 2009 Alexandrov et al. continue to cite the Englander et al. soliton hypothesis as their major motivation (Ref. [22], p. 2405): “Inspired by these ideas Englander et al. proposed that nonlinearity-induced localization of vibrational energy in dsDNA may lead to local DNA melting (DNA bubbles). This work in turn led to, the Peyrard–Bishop–Dauxois (PBD) model of DNA”.

In reality there is no confirmed experimental evidence, in the field of DNA biophysics, of any type of nonlinear dynamics. As we argue at length in Section 2.3 above, any vibrational movement of a nucleotide as a whole in water, let alone any string of nucleotides, must be severely damped making any dynamic models neglecting water inapplicable to real DNA. It is therefore not surprising that the predictions of the PBD model are in a sharp contradiction with the experimental data. In our overview of the PBD model predictions in Section 2.2 we stop short of reporting quantitative parameters of DNA fluctuations predicted by the model. This is because we have failed to find in the huge literature on the PBD model any clear-cut predictions with respect to the parameters of DNA breathing: probability of bubble formation, the average bubble length and the bubble lifetime. Still, it is obvious from the literature that the PBD model’s predictions are totally inconsistent with experimental data described in Section 3. First of all, experiment clearly testifies that there are no bubbles at all, only separate base pairs open at temperatures well below the melting range, while the bubbles predicted by the PBD model may include as many as 15 base-pairs as, for instance, Fig. 5 in [22] clearly shows. Secondly, the same figure indicates that the bubble lifetime is in the range of ps, which is also totally incompatible with the NMR data according to which the open base pair’s lifetime is in the range of $10^{-7}$ s. We conclude that the PBD model predictions contradict the experimental data on DNA described in Section 3.

Note that lately the realization of the glaring discrepancy between some predictions of the PBD model and the experimental data has been growing within the PBD community [30,75,76]. As a result, some researches try to amend the original model modifying the potential between base pairs introducing the energy barrier for base-pair closing, which increases the lifetime of the open state. In doing so, they effectively allow for some effects of water while continue to ignore other aspects of water, as discussed in Section 2.3. It must be emphasized in this connection that the microscopic partition function like in Eq. (6) is in principle flawed because it does not take into account that open (and closed) state of the base pair includes enormous number of microstates because of the flexibility of nucleotides and because of their interaction with water molecules and ions. As a result, there are only two limiting-case models,
which are consistent from the theoretical viewpoint: the full-scale MD, which is unfeasible for all kinds of reasons (see Section 2.4) and the two-state model (Section 2.6), which is inherently empirical and operates not with a Hamiltonian describing microstates like in Eq. (6) but rather with a Hamiltonian like in Eq. (17) in which each term corresponds to a multitude of microstates and presents free energies of the two generalized states and the boundaries between them. Only such an empirical two-state model automatically includes water in the form of parameters of the model. All intermediate models, between MD and the two-state model, the PBD model included, are fundamentally flawed.

Thus, our analysis shows that the PBD model is inapplicable to DNA. Therefore, all kinds of speculations (see [24] and references therein) concerning the possible role of fluctuational bubbles in DNA, predicted by the PBD model, in interaction of DNA with regulatory proteins, are groundless. As to a recent claim about involvement of the bubble dynamics in the effects of electromagnetic radiation in the THz range on the stem cells [25], in the light of our critique of the PBD model this claim looks like totally speculative.

4.2. The formaldehyde method revisited

It is quite understandable that the NMR approach to study DNA fluctuations has overshadowed the formaldehyde method: the NMR method allows to follow individual base pairs whereas the formaldehyde method can be applied only to long DNA molecules and it does not provide with the single bp resolution. Still, it is important to use an alternative method since we do not know precisely the nature of base-pair open state. Moreover, applying different approaches we may find indications of other types of DNA fluctuations, not only base pair openings. Nobody has ever tried to apply the PBD model to study theoretically the formaldehyde kinetics. The analysis of formaldehyde data was totally based on the two-state model [9]. This work was summarized in [10]. It was concluded that in addition to formaldehyde reaction with open base pairs, the so-called “outside reaction” with amino groups occurs, which does not require open base pairs. The mechanism of this “outside reaction” remained a mystery [10]. However, there were pretty convincing experimental evidences that the “outside reaction” happened and the theoretical modeling showed that without such reaction experiment and theory could not be fully reconciled. Only if one assumed that a fluctuation happened with the probability about $10^{-2}$ making amino groups (but not imino groups) fully accessible for formaldehyde, the theory quantitatively explained experiment [10]. It has also been assumed that the exchange of amino protons may proceed without full base pair opening [77].

Very recently, the issue of formaldehyde reaction with amino groups has been revisited. Adopting to DNA a sophisticated computational approach developed and verified previously to study hot spots for small molecule’s binding to proteins, Bohnuud et al. demonstrated that in the DNA B form the cytosine’s amino nitrogen is inaccessible to the formaldehyde attack [74]. Since such attack is a prerequisite for amino group hydroxymethylation, in B-DNA the “outside reaction” must not be possible. However, when the authors assumed that the base pair adjacent to the cytosine under consideration (which formed the canonical Watson–Crick, WC, base pair with guanine) flipped into the Hoogsteen (HG) base pair, the formaldehyde attack on the cytosine amino nitrogen became very much possible [74].

A transient formation of HG base pairs in DNA under normal conditions was claimed by Al-Hashimi’s group based on carbon NMR [80,81]. These pairs are supposed to form a highly unusual conformation. Indeed, HG base pairing has been well documented within triplexes but there the DNA strands connected by HG pairing have a parallel orientation (reviewed in [78]). Since in the normal DNA duplex the strands are antiparallel, the transient flipping of a WC bp into the HG bp could be possible only if the purine nucleoside in the HG pair flips from usual anti-conformation to less favorable syn conformation (see Fig. 4). Another requirement for HG pairing is protonation of cytosine [78], which additionally destabilizes HG G*C*+ pairs (we designate HG pairing by the ⋆ symbol to distinguish it from WC pairing, which we designate as -). We see why the claim by Al-Hashimi’s group about fluctuational formation of HG pairs in DNA, and with much higher probability than bp opening, was met with scepticism by some DNA experts [80,82]. Many questions need to be answered. How could such fluctuations be overlooked before in numerous proton NMR studies, which were discussed above, in Section 3? Should the HG breathing create more bendable spots in the double helix affecting the DNA persistence length, making it potentially pH-dependent? Trying to address these issues, one of us has performed a comprehensive analysis that showed that the proton NMR data do not contradict the findings by Al-Hashimi’s group [83]. With respect to the persistence length, most probably transient formation of HG pairs do not affect DNA bendability since HG pairs preserve stacking interactions, as Fig. 4 clearly shows. Still, accurate measurements of the DNA persistence length as a function of pH at neutral and especially acidic range could reveal
Fig. 4. (A) Hoogsteen (HG) A$^*$T and G$^*$C$^+$ 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.

some effect and it would be an independent evidence of the HG breathing. Nobody has done such studies yet because no effect was expected before the HG breathing claim [80,83]. Such studies are very much feasible since methods to measure the persistence length with a very high accuracy have been recently developed by Vologodskii’s group [79]. Obviously, much more studies must be done before the HG breathing is fully recognized.

Bohnud et al. have hypothesized that the “outside reaction” of formaldehyde with cytosine amino groups, suggested in early studies of formaldehyde interaction with DNA [10], proceeds through transient flipping to the HG conformation of base pairs adjacent to cytosines [74]. Thus, the full explanation of the formaldehyde reaction with DNA is achieved by invoking two types of fluctuations: full base pair openings with parameters stated in Section 3.4 and HG breathing with the probability about $10^{-2}$ [82].

5. Conclusion

Our comprehensive analysis of the extensive literature on fluctuational opening of the DNA double helix makes it possible to conclude that the basic parameters of the phenomenon remains as they were established a quarter-century ago [54,72]: only individual base pairs open at temperatures well below melting range with probability at the order of $10^{-5}$ and base pairs live open for about $10^{-7}$ s. No “bubbles” consisting of many open bases are fluctuationally formed in the linear DNA duplex at temperatures well below melting range. We also conclude that any theoretical analysis based on a microscopic partition function, which includes an energy function, or a Hamiltonian, written for a very simplified model of DNA that fully ignores water, is highly deficient. Such approaches ignore the fact that open (and closed) state of the base pair includes a huge number of microstates because of the internal flexibility of the nucleotide and because of a large number of microstates corresponding to its interaction with water molecules and ions. Only two limiting-case approaches are, in principle, consistent from the theoretical viewpoint: the full-scale MD and the empirical two-state model. In the latter model the energy function in the partition function includes free energies of the open and closed states so that the multitude of the microstates corresponding to both states is automatically allowed for. All intermediate models between MD and the two-state model, including the PBD model, are fundamentally flawed. Although MD simulations may eventually become applicable to consider this type of fluctuations, they still have a long way to go. Therefore, at present we can only rely upon the two-state model. It must be emphasized that
the two-state model allows not only explain the biophysical data on fluctuations in unstressed DNA but also makes it possible to explain DNA behavior under extreme bending of the molecule [12].

Acknowledgements

The authors thank Victor Yakhot and Alexander Vologodskii for valuable discussion and useful comments. We thank Alexander Grosberg for reading the first version of the paper and making numerous useful comments and suggestions. One of us (S.P.) acknowledges Dr. D.S. Kothari Post Doctoral Fellowship of University Grants Commission, India for financial support.

References

[16] As it is a common practice these days, we call “Hamiltonian” any energy function of the system, not necessarily the true Hamiltonian function with kinetic energy expressed via momenta of the particles.
[33] Yakhot V. Personal communication 2013.
Note that in his recent recollections, Peter von Hippel tells a very different story about early studies of DNA breathing by the formaldehyde reaction with thymine. Indeed, in the second paragraph of the right column on p. 3288, they clearly indicate: “The opening probability obtained by Frank-Kamenetskii’s group based on virtually the same experimental data, because of their unjustified neglect of the base pair opening probability claimed in a paper, which wrapped up his and McGhee’s series of four very important publications on the issue [65]. So one has to check with the original paper [65] where the opening probability was designated as $\theta$ (see Eq. (3a) in [65]) and it is claimed at the bottom of p. 3290 in [65] that the agreement between theory and experiment is achieved when $\theta = 7.1 \times 10^{-3}$. As it is argued in detail in [9,53], McGhee and von Hippel arrived to this figure, which is 3 orders of magnitude higher than the figure obtained by Frank-Kamenetskii’s group based on virtually the same experimental data, because of their unjustified neglect of the formaldehyde reaction with thymine. Indeed, in the second paragraph of the right column on p. 3288, they clearly indicate: “...we decided to focus on experimental conditions of low pH and low temperature where only adenine reaction need be considered.” Detailed computer simulations of the formaldehyde reaction with DNA performed by Frank-Kamenetskii’s group demonstrated that the thymine reaction could never be neglected, even at acidic pH (see [10,53]).
[82] The fact that GH G+·C pairs require protonation may actually support the GH breathing theory. First, Al-Hashimi’s group observed a pH-dependence of the high-energy conformation they detected in case of G–C pairs [80,81]. Secondly, although the comprehensive theoretical model, which includes base pair opening and the outside reaction, almost explained the experimentally observed pH-dependence of formaldehyde reaction with DNA, some difference remained. Indeed, when pH is increased by 1 unit (from 5.5 to 6.5), theory predicted that the characteristic rate of DNA unwinding increased by 1.9 times while the experimental figure was only 1.5 [10]. But theoretical simulations described in [10] assumed that the outside reaction was pH-independent. If fluctuational formation of HG pairs were responsible for the outside reaction, as Bohnuud et al. have suggested [74], the theoretical figure would decrease, becoming closer to the experimental value, because of a significant decrease of the probability of G+·C·HG pairs formation with increasing pH.